5. Auditory Interfaces

S. Camille Peres, University of Houston Clear Lake
Virginia Best, Boston University
Derek Brock, United States Naval Research Laboratory
Christopher Frauenberger, University of London
Thomas Hermann, Bielefeld University
John G. Neuhoff, The College of Wooster
Louise Valger∂æur Nickerson, University of London
Barbara Shinn-Cunningham, Boston University
Tony Stockman, University of London

Auditory Interfaces are bidirectional, communicative connections between two systems—typically a human user and a technical product. The side toward the machine involves machine listening, speech recognition, and dialogue systems. The side towards the human uses auditory displays. These can use speech or primarily non-speech audio to convey information. This chapter will focus primarily on the non-speech audio used to display information, although in the last section of the chapter, some intriguing previews into possible non-speech audio receptive interfaces will be presented.

Auditory displays are not new and have been used as alarms, for communication, and as feedback tools for many decades. Indeed, in the mid 1800s, an auditory display utilizing Morse code and the telegraph ushered in the field of telecommunication. As technology has improved, it has become easier to create auditory displays. Thus, the use of this technology to present information to users has become commonplace with applications ranging from computers to crosswalk signals (Brewster, 1994; Massof, 2003). This same improvement in technology has coincidentally increased the *need* for auditory displays. Some of the needs that can be met through auditory displays are: (a) presenting information to visually-impaired people, (b) providing an additional information channel for people whose eyes are busy attending to a different task, (c) alerting people to error or emergency states of a system, and (d) providing information via devices with small screens such as PDAs or cell phones that have a limited ability to display visual information. Furthermore, the auditory system is well suited to detect and interpret multiple sources and types of information as will be described in the chapter.

Audio displays (or auditory interfaces) as they are experienced now, in the beginning of the 21st century, are more sophisticated and diverse than the bells and clicks of the past. As mentioned previously, this is primarily due to the increasing availability of powerful technical tools for creating these displays. However, these tools are just recently accessible to most engineers and designers and thus, the field of auditory displays is somewhat in its adolescence and for many is considered relatively new. Nevertheless, there is a substantial and growing body of work regarding all aspects of the auditory interface. Because this field is young and currently experiencing exponential growth, the lexicon and taxonomy for the different types of auditory displays is still in development. A discussion of the debates and nuances regarding the development of this taxonomy is

not appropriate for this chapter, but the interested reader will find more regarding this in Gregory Kramer's Book on Auditory Displays (1994), from the International Community on Auditory Displays (ICAD; www.icad.org), as well as other sonification sites (e.g., http://sonification.de).

For the benefit of the reader, the terms used in this chapter are outlined and defined below. We have organized the types of auditory displays primarily by the method or technique used to create the display. Additionally, all of the sound examples are available on the website www.beyondthegui.com.

Sonification of complex data

Sonification is the use of non-speech sound to render data, either to enhance the visual display or as a purely audio display. Sonification is routinely used in hospitals to keep track of physiological variables such as those from electrocardiogram (ECG) machines. Audio output can draw the attention of medical staff to significant changes while they attend other patients. Other sonifications include the rhythms in electroencephalogram (EEG) signals that can assist with the prediction and avoidance of seizures (Baier, Hermann, Sahle, and Ritter, 2006) and sonification of the execution of computer programs (Berman and Gallagher, 2006). There are different types of sonification techniques, and these will be elaborated throughout the chapter and particularly in section 5.2.4.

Considerable research has been done regarding questions that arise in designing effective mappings for sonifications of data. Among the key issues are *voicing*, *property*, *polarity*, and *scaling/context*. Voicing deals with the mapping of data to the sound domain, i.e., given a set of instruments, which one should be associated with which variable? For example, Quinn[1] sonified stock prices using instruments representative of where the stocks originated in the world e.g. Asian instruments indicate Asian stocks. Property deals with how changes in a variable should be represented, e.g., should changes in a data variable be mapped to pitch, amplitude or tempo? Polarity deals with the way a property should be changed (Walker, 2002), i.e., should a change in a variable cause the associated sound property to rise or fall? Suppose weight is mapped to tempo: should an increase in weight lead to an increase in tempo, or a decrease, given that increasing weight generally would be associated with a slower response? Scaling deals with how quickly a sound property should change. For instance, how can we convey an understanding of the absolute values being displayed: the maximum and minimum values being displayed, where the starting value is in that scale, when the data crosses zero? How much of a change in an original data variable is indicated by a given change in the corresponding auditory display parameter?

Audification

A very basic type of auditory display, called *audification* is simply presenting raw data using sound. This will be described in more detail in section 5.2.4, but essentially, everything from very low frequency seismic data (Hayward, 1994) to very high

[1] http://www.quinnarts.com/srl/

frequency radio telescope data (Terenzi, 1988) can be transformed into perceptible sound. Listeners can often derive meaningful information from the audification.

Symbolic/semantic representations of information

Similar to other types of interfaces, it is often the case that auditory interfaces or displays are needed for a task other than data analysis or perceptualization. For instance, GUIs use icons to represent different software programs or functions within a program. This type of visual display is more semantic in nature and does not require analysis on the part of the user.

The auditory equivalents of icons are auditory displays known as auditory icons and earcons. These displays are very useful in translating symbolic visual artifacts to auditory artifacts and will be discussed several times through the chapter. An example of an *auditory icon* is the paper "crinkling" noise that is displayed when a user empties the "Trash" folder. This technique employs sounds that have a direct and thus intuitive connection between the auditory icon and the function or item.

Earcons are a technique of representing functions or items with more abstract and symbolic sounds. For example, just as Morse code sound patterns have an arbitrary connection to the meaning of the letter they represent, the meaning of an Earcon must be learned. These types of displays can be particularly appropriate for programs that have a hierarchical structure as they allow for the communication of the structure in addition to the representation of the functions.

A new symbolic/semantic technique for auditory displays that shows some promise is the use of *spearcons*. These are non-speech cues used in the way that icons or earcons would be. They are created by speeding up a spoken phrase until it is not recognized as speech (Walker, Nance, and Lindsay, 2006). This representation of the spoken phrase, for example someone's name in a phone list, can be slowed down to a recognizable level to facilitate learning the association between the spearcon and the name. Once this association is made, the spearcon can be played using the shortest duration to reduce the amount of time necessary to present the auditory display.

It is very important to understand, and should be clear after reading the entire chapter, that these techniques are not normally and sometimes not even ideally used exclusively. Furthermore, they are not necessarily independent of each other. For instance, in daily weather sonifications (Hermann, Drees, and Ritter, 2005) most data variables were displayed via parameterized auditory icons (e.g. 'water sounds' gave an iconic link to rain, the duration of the sound allowed the user to judge the amount of rain per unit time). When designing and testing auditory displays, designers can consider three different dimensions or axes: the interpretation level—from analogic to symbolic; interactivity—from non-interactive to tightly closed interaction; and "hybridness"—from isolated techniques to complex mixtures of different techniques. The ability to use these three dimensions gives the designer a wide and intriguing range of tools to better meet the needs of the users with auditory interfaces.

5.1 Nature of the Interface

5.1.1 Basic Properties of Sound

Sound arises from variations in air pressure caused by the motion or vibration of an object. Sounds are often described as pressure variations as a function of time, plotted as a waveform. Figure 5.1 shows the waveform for a pure sinusoid, which is periodic (repeats the same pattern over and over in time) and is characterized by its frequency (number of repetitions per second), amplitude (size of the pressure variation around the mean, and phase (how the waveform is aligned in time relative to a reference point). All complex sounds can be described as the sum of a specific set of sinusoids with different frequencies, amplitudes, and phases (or "Fourier analysis"). Sinusoids are often a natural way to represent the sounds we hear because the mechanical properties of the cochlea break down input sounds into the components at different frequencies of vibration. Any incoming sound is decomposed into its component frequencies, represented as activity at specific points along the cochlea. Many natural sounds are periodic (such as speech or music) and contain energy at a number of discrete frequencies that are multiples of a common (fundamental) frequency. Others are non-periodic (such as a clicks or a white noise) and contain energy that is more evenly distributed across frequency.

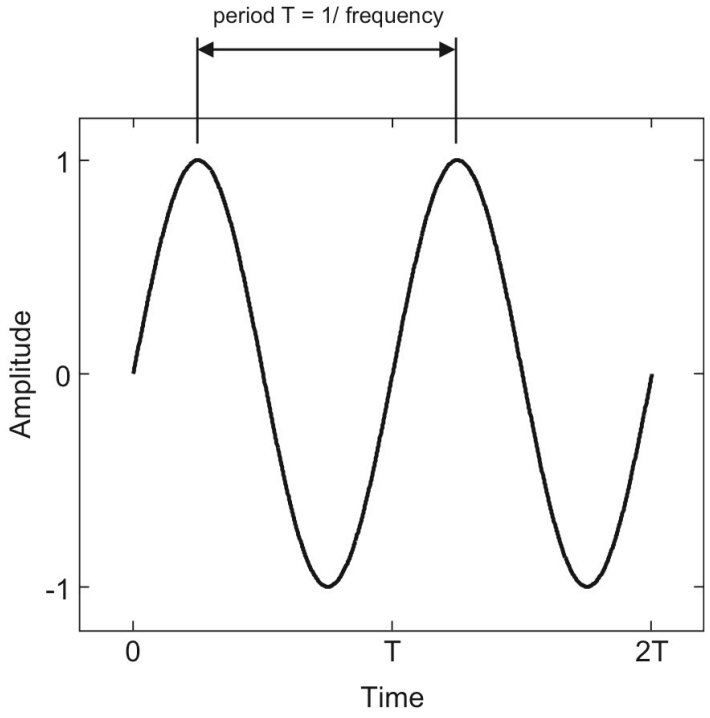

Figure 5.1. The waveform for a pure sinusoid. The elements of a waveform are: frequency (number of repetitions per second), amplitude (size of the pressure variation around the mean), and phase (how the waveform is aligned in time relative to a reference point)

5.1.2 Human Sensitivity to Auditory Dimensions

These basic properties of sounds give rise to a number of perceptual dimensions that form the basis of auditory displays. The effectiveness of an auditory display depends critically on how sensitive listeners are to changes along the dimensions used to represent the relevant information. Below is a brief examination of human sensitivity to the dimensions of frequency and pitch, loudness, timbre, and spatial location (for an extensive discussion see Moore, 2003).

Frequency and Pitch

As mentioned above, sound frequency is a fundamental organizing feature of the auditory system and a natural dimension for carrying information in auditory displays. Effective use of the frequency domain must take into account both the range of frequencies to which human ears respond and how well listeners can distinguish neighboring frequencies within this range.

Human listeners are able to detect sounds with frequencies between about 16 Hz and 20 kHz, with sensitivity falling off at the edges of this range. Frequency resolution (i.e., the ability to distinguish different frequencies) within the audible range is determined by the peripheral auditory system, which acts like a series of overlapping filters. Each filter is responsive to the sound energy in a narrow range of frequencies and has a bandwidth that varies with frequency. For frequencies below 200 Hz, the bandwidth is constant at around 90 Hz, while for higher frequencies it is approximately 20% of the center frequency. The width of these filters, or "critical bands" determines the minimum frequency spacing between two sounds required to perceive them separately. Sounds that fall within the same "critical band" will generally be difficult to separate perceptually.

Pitch is the subjective attribute of periodic sound that allows it to be ordered on a musical scale or to contribute to a melody. For pure sinusoids, the pitch is monotonically related to the frequency, with higher frequencies giving rise to higher pitches. For complex, periodic sounds, pitch is monotonically related to the fundamental frequency. Most musical instruments produce periodic sounds that have a strong pitch. Periodic sounds without a strong pitch can nonetheless be told apart (discriminated) based on their spectral content.

Loudness

The subjective experience of loudness is related to its physical correlate, the intensity of a sound waveform (or square of the pressure amplitude; Figure 5.1), and has been described using several different scales (such as the sone scale, Stevens, 1957). Human listeners are sensitive to a large range of intensities, with sensitivity that is roughly logarithmic so that the smallest detectable change in loudness for wideband sounds is approximately a constant fraction of the reference loudness. Listeners can detect intensity changes of just a few decibels (dB, defined as 10 times the logarithm of the ratio of the intensities) for many types of stimuli.

Timbre

For a complex sound, energy can be distributed in different ways across frequency and time, giving the sound its quality or timbre. For example, two complex tones of the same pitch and loudness may differ in their timbre due to the detailed structure of the waveform (e.g. differences in the relative magnitudes of the different frequency components). In intuitive terms, these differences are what distinguish two different instruments (e.g. the bowed violin and the percussive piano) playing the same note or chord from the same location, with the same loudness.

Temporal Structure

Most sounds are not stationary, but turn on and off or fluctuate across time. Human listeners are exquisitely sensitive to such temporal changes. Temporal resolution is often quantified as the smallest detectable silent gap in a stimulus, and is on the order of a few milliseconds for human listeners (Plomp, 1964). Temporal resolution can also be described in terms of how well listeners can detect fluctuations in the intensity of a sound over time (amplitude modulation). Modulation detection thresholds are constant for rates up to about 16 Hz, but sensitivity decreases for rates from 16 - 1000 Hz, where modulation can no longer be detected. For human listeners, relatively slow temporal modulations are particularly important as they contribute significantly to the intelligibility of naturally spoken speech (Shannon, Zeng, Kamath, Wygonski, and Ekelid, 1995).

Spatial Location

Sounds can arise from different locations relative to the listener. Localization of sound sources is possible due to a number of physical cues available at the two ears (see Carlile, 1996 for review). Differences in the arrival time and intensity of a sound at the two ears (caused by the head acting as an acoustic obstacle) allow an estimation of location in the horizontal plane. For example, a sound originating from the left side of a listener will arrive at the left ear slightly earlier than the right (by tens to hundreds of ms) and will be more intense in the left ear than in the right (by up to tens of dB). In addition, the physical structure of the outer ear alters incoming signals and changes the relative amount of energy reaching the ear at each frequency. This spectral filtering depends on the direction of the source relative to the listener, and thus provides directional information to complement that provided by the interaural cues (e.g. allowing elevation and front-back discrimination). For estimating the distance of sound sources, listeners use a number of cues including loudness, frequency content, and (when listening in an enclosed space) the ratio of the direct sound to the energy reflected from nearby surfaces such as walls and floors (Bronkhorst and Houtgast, 1999).

The spatial resolution of the human auditory system is poor compared to that of the visual system. For pairs of sound sources presented in succession, human listeners are just able to detect changes in angular location of around 1° for sources located in the front, but require changes of 10° or more for discrimination of sources to the side (Mills, 1958). For the case of simultaneous sources, localization is well preserved as long as the sources have different acoustic structures and form clearly distinct objects (Best, Gallun, Carlile, and Shinn-Cunningham, 2007).

5.1.3 Using Auditory Dimensions

When using sound to display information, the available auditory dimensions and human sensitivity to these dimensions (reviewed above) are critical factors. However, there are also a number of design questions related to how to map data to these dimensions in an effective way.

As an example, it is not always clear how data "polarity" should be mapped. Intuitively, increases in the value of a data dimension seem as though they should be represented by increases in an acoustic dimension. Indeed, many sonification examples have taken this approach. For example, in the sonification of historical weather data, daily temperature has been mapped to pitch using this "positive polarity," where high pitches represent high temperatures and low pitches represent low temperatures (Flowers, Whitwer, Grafel, and Kotan, 2001). On the other hand, a "negative polarity" is most natural when sonifying size, whereby decreasing size is best represented by *increasing* pitch (Walker, 2002). To add to the complexity of decisions regarding polarity, in some cases individual listeners vary considerably in their preferred polarities (Walker and Lane, 2001).

As another example, redundant mappings can sometimes increase the effectiveness with which information is conveyed. Recent work (Peres and Lane, 2005) has shown that the use of pitch and loudness in conjunction when sonifying a simple data set can lead to better performance, but other conjunctions may not.

5.1.4 Perceptual Considerations with Complex Displays

Multiple Mappings

With multidimensional data sets, it may be desirable to map different data dimensions to different perceptual dimensions. As an example, the pitch and loudness of a tone can be manipulated to simultaneously represent two different parameters in the information space (see Pollack and Ficks, 1954). However, recent work has shown that perceptual dimensions (such as pitch and loudness) can interact such that changes in one dimension influence the perception of changes in the other (Neuhoff, 2004).

In many cases, the most effective way of presenting multiple data sets may be to map them to auditory objects with distinct identities and distinct spatial locations. These objects can be defined on the basis of their identity (e.g. a high tone and a low tone) or their location (e.g. a source to the left and a source to the right). This approach theoretically allows an unlimited number of sources to be presented, and offers the listener a natural, intuitive way of listening to the data.

Masking

Masking describes a reduction in audibility of one sound caused by the presence of another. A classic example of this is the reduction in intelligibility when speech is presented against a background of noise.

In auditory displays with spatial capabilities, separating sources of interest from sources of noise can reduce masking. For example, speech presented against a background of noise is easier to understand when the speech and the noise are located in different places (e.g., one on the left and one on the right). In such a situation, the auditory system is able to use differences between the signals at the ears to enhance the perception of one source. In particular, it is able to make use of the fact that one of the two ears (the one nearest the speech target) is biased acoustically in favor of the target sound due to the shadowing of the noise by the head (Bronkhorst, 2000).

Auditory Scene Analysis and Attention

Another crucial consideration when delivering multiple signals to a listener is how the auditory system organizes information into perceptual "streams" or "objects." A basic principle of auditory scene analysis (Bregman, 1990) is that the auditory system uses simple rules to group acoustic elements into streams, where the elements in a stream are likely to have come from the same object. For example, sounds that have the same frequency content or are related harmonically are likely to be grouped into the same perceptual stream. Similarly, sounds that have synchronous onsets and offsets and common amplitude and frequency modulations are likely to be grouped together into one perceptual object. In addition, sounds that are perceived as evolving over time from the same spatial location tend to be perceived as a related stream of events. Grouping rules can be used in auditory display design when it is desirable that different signals be perceived as a coherent stream, but unwanted grouping can lead to disruptions in the processing of individual signals.

Confusion about which pieces of the acoustic mixture belong to which sound source are quite common in auditory scenes containing sounds that are similar along any of these dimensions (Kidd, Mason, and Arbogast, 2002). Related to this issue, sources in a mixture can compete for attention if each source is particularly salient or contains features that may be relevant to the listener's behavioural goals. By making a target source distinct along one of the perceptual dimensions discussed above (e.g. by giving it a distinct pitch or spatial location), confusion can be reduced as the listener's attention will be selectively directed along that dimension. For example, when a listener must attend to one voice in a mixture of competing voices, the task is much easier (and less confusions are made) when the target voice differs in gender from its competitors (Darwin and Hukin, 2000; Shinn-Cunningham, 2005).

Auditory Memory

In complex auditory displays, the capabilities and limitations of auditory memory are important considerations. The auditory system contains a brief auditory store ("immediate" or "echoic" memory) where a crude representation of the sensory stimulus is maintained, normally for no longer than 2 seconds (Neisser, 1967). This store makes it possible to retain a sound temporarily in order to make comparisons with later-arriving sounds, as well as to process simultaneous stimuli in a serial fashion (Broadbent, 1958). When stimuli are processed in more detail (such as the semantic processing of speech, or the learning of a sound pattern in the environment), there is the possibility for more permanent, categorical representations and long-term storage.

5.2 Technology of the Interface

5.2.1 Auditory Display Systems

This section provides a brief overview of the technology required to create auditory displays. A typical auditory display system encompasses the various components sketched in Figure 5.2: (A) data representation, (B) the main application (processing loop), which uses the data to determine when sounds should be created, (C) auditory display techniques to render an acoustic signal (digital audio signal) based on the data, and finally, (D) technical sound display systems such as sound cards, mixers, amplifiers, headphones or loudspeakers to convert the digital audio signals to audible vibrations at the user's eardrums. These components establish a closed-loop system that integrates the user and can be recreated in almost any auditory interface. Often, it is possible for the user to interact with the system—this can be a very simple interaction like starting the sonification playback or can involve more complex continuous interactions, which are the particular focus of Interactive Sonification (described later).

The loudspeakers or "physical" displays are the only visible (and audible) part in this chain and are often referred to as the "front end" of the auditory display system (ADS). Most of the technology for the ADS is hidden behind the curtain of computation, algorithms and signal processing. The current section will focus mostly on these "back end" or invisible parts. For further reading on the sound engineering and hardware of ADSs, the reader should consult specialized literature such as Miranda, 1998.

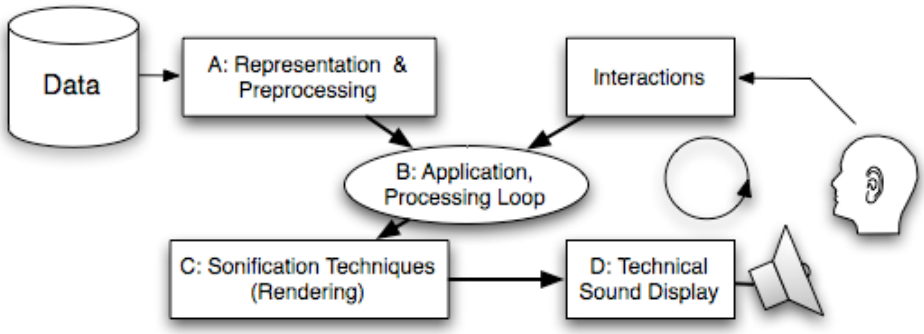

Figure 5.2. Sketch of information flow in a typical auditory display system.

Concerning the sound hardware (or front end of the ADS), the choice of whether to use headphones or speakers is determined basically by issues such as: (a) privacy, (b) mobility, (c) practicality, (d) isolation and/or (e) user goals. Loudspeakers, for instance do not require the listener to wear any electronic equipment and thus increase the user's mobility, yet their sounds are audible for everyone in the room. Headphones, in contrast, allow a personal auditory display, yet may be impractical since they may interfere with auditory perception of the natural surrounding and thus isolate the user from his/her surroundings.

Loudspeaker and headphone systems also differ in their ability to communicate the source location of sound. This aspect of auditory interfaces is highly relevant for applications where the goal is to direct the user's focus of attention. Multi-speaker

systems (Pulkki, 1997) are well suited if the user can be assumed to be stationary and located in a "sweet spot." Typically headphones lead to the user perceiving the source within his or her head, however, 3D-spatialization with headphones can be achieved by modeling the filter effect of the outer ear (see Spatial Location in section 5.1.2). This is done mathematically with an adjustment of the source sound using Head-Related Transfer Functions (HRTF; Carlile, 1996). To be convincing, however, the head position and orientation of the listener have to be tracked so that the perceived sound source position can be kept constant while the user moves his or her head.

The technology for the front end of the ADS (D in Figure 5.2) is well developed and continues to advance due to the work of sound and product engineers. However, the larger back end of the ADS, consisting of the technologies to compute sound signals for auditory displays (A-C in Figure 5.2), is much younger and thus not as well developed. This back end will be the focus of sections 5.2.2 through 5.2.4. A note for the human factors/usability practitioner: In order to describe the technology of the auditory interface, the information that will be displayed to the user must often be described in terms of programming code or computer science. Thus, for the benefit of those readers wanting to create auditory displays, in these technology-oriented sections, we will use mathematical descriptions and program code examples. However, those less familiar with programming may want to skim over the mathematics/programming details.

5.2.2 Data Representation and Processing

Let us start the discussion of technologies from the point in the information circle where information is created, measured or becomes available within a computer system, (depicted in Figure 5.2 as A). Information may have highly different appearance, from a symbolic (textual) level (e.g. an alarm condition that a cooling system failed) to a more analogic level of raw measurements (e.g. temperature values). Quite often, data are organized as a table of numbers, each column representing a different feature variable, each row representing measurements for a single record. In a census data set, for instance, columns could be features like 'income', 'weight', 'gender', while rows could represent different persons. We call such a representation a data set X, and in most cases data or information can be recoded into such a form. We will refer to rows of X as $\vec{x}^{\alpha}$, and to the ith feature as x_i^{α}. Many auditory displays—ranging from applications in process monitoring, exploratory data analysis sonifications, to table viewers for blind users—operate using this type of data representation. Another frequent case for auditory displays is the communication of single events (e.g., to signal that an e-mail arrived). The message can be characterized by a set of feature values (e.g. sender, email length, urgency, existing email attachments, etc). These values form a row vector $\vec{x}^{\alpha}$ following the above representation. An auditory display technique should be able to represent all possible feature vectors using the systematic transformation of event data to sound.

Statistics and data mining (Fayyad, Piatetsky-Shapiro, Smyth, and Uthurusamy, 1996) are the disciplines concerned with explaining all of the peculiarities of how data is structured and summarized and a thorough introduction would exceed the scope of this chapter. Central aspects of data, though, are the range (minimum/maximum values), whether the data are discrete or continuous, and whether a variable is nominal or ordinal. It is important to identify these aspects of data because certain auditory variables are often considered better suited to represent certain data types, e.g. timbre matches better to

nominal variables whereas frequency fits well to a feature variable with given order. Data features with a zero value (e.g. velocity of a car) match with acoustic variables that have a zero (e.g., loudness, vibrato frequency, pulsing rate, etc.).

5.2.3 Sound Synthesis

Sound synthesis is the technological basis for controlling sound characteristics by data. In rare cases, it might be possible to simply record sounds for every possible condition or event. But whenever full control over all sound characteristics is wished or needed, sound synthesis is essential.

Digital sound signals are vectors of numbers that describe the sound pressure at every moment. Real-time sound computing is thus computationally quite demanding and scales with the number of independent audio channels. There are some powerful programming systems for sound synthesis. The code examples below are given for SuperCollider (McCartney, 1996) a versatile, compact, powerful and open-source textual programming system. Pure Data is another graphical engines (also cross-platform and open-source; Puckette, 1997).

Additive/Subtractive Synthesis

Additive Synthesis is the creation of complex sounds from simple ingredients. These ingredients are the building blocks in a bottom-up approach. The building blocks are simple signals (such as sine waves $b_\omega(t) = \sin(\omega t + \varphi)$) and their superposition represents the result of additive sound synthesis:

$$s(t) = w(t) \cdot \sum_{i=1}^{N} a_k b_{\omega_k}(t) = w(t) \cdot \sum_{i=1}^{N} a_k \sin(\omega_k t + \varphi_k) \qquad (5.1)$$

To obtain harmonic timbres, frequencies ω_k are chosen as integer multiples of a fundamental frequency ω_1. The coefficients a_i determine how strong each component contributes to the sound. An example of achieving additive synthesis in SuperCollider, is: for only two partial tones of 440 Hz the designer would use

```
{SinOsc.ar(440, mul: 0.4) + SinOsc.ar(880, mul: 0.2)}.play.
```

Sound example S1[2] provides 1 sec of the sound. Sound examples S2 – S4 demonstrate some typical additive synthesis sounds.

Subtractive Synthesis takes the opposite (top-down) approach and creates complex timbres by removing material from a spectrally rich source such as sawtooth shaped signals (refer to Section 5.1 for more on signal shapes), pulse trains, or spectrally rich noise signals. A good introduction to subtractive synthesis can be found in (Moore, 1990). Sound examples S5 – S8 demonstrate some typical sounds of subtractive synthesis and sound examples S9 - S12 demonstrate different filters.

[2] all sound examples also available at http://sonification.de/publications/BeyondGUI/

Wavetable Synthesis

One of the most practically relevant synthesis techniques is wavetable synthesis where basically a recorded version of a real-world sound is used to generate the synthesized output. Mixing and manipulating the synthesis algorithm allows the audio designer to create novel sounds or to play a sample at different musical notes. Many commercial sound synthesizers rely on this technique, and most of the auditory icons (discussed later) are produced using wavetable synthesis.

In SuperCollider, a wavetable is represented by a `Buffer`, and a `PlayBuf` unit generator can be used to play the buffer at arbitrary speed, as demonstrated in the following code example (sound example S13), where the playback rate is slowly modulated by the sine oscillator:

```
b = Buffer.read(s,"soundsample.wav") ;

{PlayBuf.ar(1, b.bufnum, SinOsc.kr(0.2, mul: 0.4, add:
1)*BufRateScale.kr(b.bufnum), loop: 1)}.play
```

Other Synthesis Techniques

Other synthesis techniques include Granular Synthesis, Physical Modeling sound synthesis, FM-Synthesis, and Nonlinear Synthesis. The discussion of these techniques easily fills books, and the interested reader should look to Moore (1990), Roads (2001), and Cook (2002) for more information.

5.2.4 Auditory Display Techniques in a Nutshell

This section focuses on different auditory display techniques. In general, these techniques are algorithms that connect the data that will be displayed to sound synthesis techniques (described in 5.2.3). As mentioned before, auditory display techniques can roughly be characterized as symbolic or analogic. We start here with the symbolic sonification techniques.

Auditory Icons: Auditory Icons, as mentioned in the introduction, represent specific messages via an acoustic event that should enable the quick and effortless identification and interpretation of the signal with respect to the underlying information. These sounds need to be either selected from a database of recordings, or synthesized according to the data features (in the example: file size), which is practically achieved by adapting appropriate sound synthesis algorithms (see 5.2.3).

Earcons: Different from Auditory Icons, Earcons inherit structural properties from language as a more abstract and highly symbolic form of communication (Blattner, Papp, and Glinert, 1994). These sounds can be built using concatenation, which allows the designer to compose more complex messages from simple building blocks.

Audification: In audification, the data "speak for themselves" by using every data value as a sound sample in a sound signal $s(t)$. Since only variations above 50 Hz are acoustically perceptible (see Section 5.1), audifications often consume thousands of samples per second. The technique is thus only suited, if (a) enough data are available,

(b) data can be organized in a canonical fashion (e.g. time-indexed measurements), and (c) data values exhibit variations in the selected feature variable. Mathematically, audification can be formalized as the creation of a smooth interpolation function going through a sample of (time, value) pairs $\left(t^{\alpha}, x^{\alpha}\right)$ for all data items α. The simplest implementation of audification, however, is just to use the measurements directly as values in the digital sound signal by setting $s[n] = x^{n}$. Some sound examples for audifications demonstrate the typical acoustic result (S14[3], S15). S14 is an audification of EEG measurements, i.e., one electrode measuring the brain activity of a beginning epileptic attack (roughly in the middle of the sound example). S15 plays the same data at lower time compression. Clearly the pitch drops below the well audible frequency range and the epileptic rhythm is perceived as audible rhythm of events.

Parameter Mapping Sonification: Parameter mapping sonification (PMS) is the most widely used sonification technique for generating an auditory representation of data. Conceptually, the technique is related to scatter plotting, where different features of a data set determine different graphical features of symbols (such as x-position, y-position, color, size, etc.) and the overall display is a result of the superposition of these graphical elements. To give an example, imagine a data set of measurements for 150 iris flowers. For each flower, measurements of the petal length, sepal length, petal width and sepal width are listed. A parameter mapping sonification (S16) could for instance map the petal length to the onset time of sonic events, the sepal length to pitch of sonic events, the petal width to brilliance and the sepal width to duration. The resulting sonification would allow the listener to perceive how the data are organized in time or change with time. Each sound event represents a single flower, while the PMS displays the entire data set of measurements of all 150 flowers!

Model-based Sonification: A structurally very different approach to PMS is Model-based Sonification (MBS; Hermann, 2002). In PMS, data directly control acoustic attributes, however in MBS the data are used to create a sound-capable dynamic model. The result of this is that a sonification model will not sound at all unless excited by the user and thus puts interaction into the fore. The set of rules regarding how to create a virtual sound object from data is called a sonification model and they can be designed in a task-oriented way. For example, imagine that every SMS in a mobile phone is like a marble in a box. By shaking the phone the marbles would move, interact, and thereby create an acoustic response from that you would be able to infer how many text messages, of what size, etc. have arrived (see Williamson, Murray-Smith, and Hughes, 2007). The excitation here is "shaking," the dynamics are the physical laws that describe the marble motion and interactions, etc. The sonification model simulates the whole physical process and thus creates an interactive and informative sonification.

Interactive Sonification: Interactive Sonification is a special focus in auditory interfaces and can be used with many types of sonification techniques. Often there is a particular benefit results from tightly closed interaction loops between the user and a sonification system (Hunt and Hermann, 2004). All sonification techniques can be modified to be more interactive, e.g. for audification, interactive sonification can enable the user to actively navigate the data while generating the sound, etc. The rationale behind interactive sonification is that people typically get acoustic responses latency-free

[3] all sound examples also available at http://sonification.de/publications/BeyondGUI/.

as by-products of their interaction activity, and they use the acoustic feedback continuously to refine their activity, be it within a search, scan, discovery or any other task.

5.3 Current Implementations of the Interface

5.3.1 A Bit of History

Sound often helps direct our focus and describes what is going on. Listening to a car engine or the spinning of a hard-drive can offer vital clues whether the car is in good mechanical condition or the computer is finished saving a file. In early computing such incidental sounds were often used, for example beeps were introduced to indicate errors in the program or the beginning of a new iteration of a program loop. There is a long tradition of audio indicating warnings and alerts. Sound has the potential to be used in many more sophisticated ways, ranging from short sounds to indicate specific events to fully immersive spatial sound environments.

5.3.2 Why Sound is Used

In addition to those mentioned in the introduction of this chapter, there are numerous reasons why sound may be used in an interface. A very common one is in order to reinforce a visual message, such as an alert. Other reasons are outlined below:

Reducing visual overload: Visual interfaces tend to be busy, filling as much of the screen as possible with information. Constantly changing visual information can be distracting and limit the amount of information that reaches the user. Where applicable, the cognitive load can be lessened by channelling information to the ears (Brown, Newsome, and Glinert, 1989).

Reinforcing visual messages: Sending the same information to more than one sense can ensure that the user receives the information, making the interface more effective. For example, when entering a personal identification number (PIN) at an automated teller machine (ATM), the machine beeps for each number entered and the visual interface uses an asterisk to represent each number. This dual feedback can reassure the users that their input was received.

When eyes are elsewhere: Since sound can be perceived from all directions, it is ideal for providing information when the eyes are otherwise occupied. This could where someone's visual attention should be entirely devoted to a specific task such as driving or a surgeon operating on a patient (Recarte and Nunes, 2003).

When audio is more informative: Humans are very good at hearing patterns in sound. This means that at times, it is easier to understand information when it is sonified (Bly, 1982). Two prime examples of this are seismic data (Hayward, 1994) and medical monitoring data (Baier and Hermann, 2004). Users can very quickly notice a change, which may not have been as easily noticed by looking at numbers or a graph.

Small or no visual display: Unlike visual displays where the size of the interface is determined by the size of the device, audio is limited only by the sound quality that the

device can provide. It is therefore a good candidate for augmenting or replacing a visual interface.

Conveying emotion: The aesthetics of sound can have a great impact on the user's impression of an application. This is particularly obvious in video games where the sound design is carefully orchestrated to make players enjoy the game and to impact the amount of tension the player experiences.

5.3.3 Drawbacks to Using Sound

Some of the major disadvantages to using sound are annoyance and privacy. There are also dangers to using too much sound. If sound is poorly designed or used at the wrong time, users are very quick to turn it off. Some of the drawbacks are outlined below:

Annoyance: Sound is very good at drawing the user's attention. However, the urgency of the sound should be relative to the importance of the information. Obtrusive sounds for something of low importance can quickly annoy the user.

Privacy: Sound is omni-directional and thus cannot be directed at a single user unless they are using a headset or a handset. Therefore, if headsets are not employed, an auditory interface can publicize what a user is doing in an undesirable manner.

Auditory overload: Displaying too much information in sound can result in that information losing meaning and being interpreted as noise.

Interference/masking: As mentioned in 5.1.3, sounds can interfere or mask one another. Like auditory overload, this can result in loss of information. Environmental sounds can also cause interference and/or masking.

Low resolution: With visual displays, objects can be located very precisely on the screen; with spatial sound interfaces, there is a greater area that can be used but the resolution is lower (See section 5.1.2). This difference is evident, for example, in visual and audio games. In a visual game, it is possible to pinpoint a target with great accuracy on a computer screen. In an auditory game however, although the auditory display may be presented in an area much larger than a computer screen, the lower resolution that hearing affords in locating sources of audio greatly reduces the accuracy with which audio sources can be located.

Impermanence: Unlike with visual displays, where the information on the screen remains, audio is serial and once played is easily forgotten (See section 5.1.).

Lack of familiarity: Sounds take a while to get to know and can be initially confusing to a user until they know what they are listening to. As with visual icons, the sounds in an audio interface need to be easy to learn.

5.3.4 Advanced Audio Interfaces

The availability of increasingly sophisticated audio hardware and software has provided the possibility for more widespread use of audio in interfaces. This is both in terms of using audio to support richer human-computer interactions and in terms of the quality of the audio that can be used, which can be the equivalent of that used in a film or radio production. This more sophisticated use of audio may of course be as part of a multi-

modal interface, or be the basis of an audio-only display. Although much of this potential remains under exploited, this subsection will examine several application areas where these improved audio capabilities have been used to good effect.

Audio for monitoring

Audio is a well-known method for monitoring in specialist environments such as hospitals and environmental monitoring facilities such as weather and seismic activity. Other applications are stock market and network monitoring.

Accentus[4] makes Sonify!, a product for traders to (among other things) monitor stock price changes and stock progress towards a target (Janata and Childs, 2004). Sonify! uses real instruments and short melodies to create pleasant tones that are informative yet remain unobtrusive. The use of sound is particularly appropriate as traders are usually surrounded by numerous screens and work in a very visually intensive environment. They are often required to monitor data while talking on the phone. The non-speech audio can reduce the visual overload and can be listened to in the background of other auditory events like telephone conversations.

Numerous programs that use sound to monitor network traffic have been developed. The Sheridan Institute in Canada has developed *i*SIC (Farkas, 2006), which uses mathematical equations on network traffic data to create music. The Sound of Traffic (Weir, 2005) uses MIDI sounds to track traffic to specific ports on a computer. These network monitors allow administrators to hear real-time data about their sites and networks without having to purposely go to an interface or log to check activity; they allow the monitoring to be a background process. While these programs are not commercially available, they show a growing trend in the use of audio in a variety of different applications.

Audio in games

Audio plays a very important role in video games, using music to set the mood and sound effects to bring realism to the action. Audio in most modern games is a multi-layered production used to render complex auditory scenes, creating an immersive environment. Sound usually takes the form of effects such as one would hear in film – known as Foley effects[5] –and human dialog to imbue scenes with realism. This audio is increasingly synthesized dynamically rather than pre-recorded. A particular problem faced by game designers is the fact that, while broached early on in the design phase, often sounds are only incorporated into the game after the whole of the interactive graphic design has been implemented. Thus, proper sound testing is done at a very late stage. The end result is that the sound design is mostly æsthetic as opposed to informative and not nearly as powerful as it could be.

---

[4] http://www.accentus.com/

[5] Jack Foley pioneered the craft of creating sound effects using all kinds of materials to imitate sounds for films in the early days of the "talkies" for Universal Studios

Audio-only games

Audio-only games are often, but not exclusively, targeted at the visually impaired. Evolving from mostly speech and simple sound effects, today's audio games, such as Shades Of Doom (first person shooter) and Lone Wolf (submarine simulator), use sophisticated sound, transitioning smoothly between multi-layered and immersive soundscapes. Many of today's audio-only games also provide a visual interface for collaborative game play between visually impaired and sighted gamers. A typical issue audio game designers face is supporting orientation and navigation in a 3D world (Andreson, 2002), e.g., which way is the player facing, where are the walls/entrances/exits, and what is the state of other players/objects? The AudioGames web site[6] is dedicated to blind-accessible computer games and includes issues of the audio gaming magazine *Audyssey* (2006).

5.3.5 Applications of auditory interfaces to accessibility

Accessibility and the desktop

Non-speech sound has significant potential for improving accessibility, either on its own or in complement to other media, particularly for users with visual impairments, due to its ability to convey large amounts of information rapidly and in parallel. Imagine a visually impaired teacher monitoring the progress of students taking an electronic multi-choice test. The teacher could get a rapid "auditory glance" of approximately how far through the test students were and the proportion of correct answers, individually and as a whole, using the concepts described in the section on data sonification.

Visually impaired users use software called a *screen reader*, which uses synthetic speech to speak text such as menus, the state of GUI widgets and text in documents. The most popular screen readers used worldwide are Jaws® for Windows (JFW) by Freedom Scientific (2005) and Window-Eyes by GW Micro (2006). The use of non-speech audio has seen relatively little commercial uptake. The release of JFW version 5, in 2003, represented the first significant use of non-speech sound: the ability to customize feedback. These customizations, called "behaviors", are defined in schemes. Each scheme can be associated with a specific application. Examples include: to identify focus on a particular type of widget, the state of a check box, upper/lower case, the degree of indentation, and the values of HTML attributes.

The inclusion of non-speech sound is intended to improve efficiency and effectiveness, such as in navigation of the GUI or screen-based proof reading of lengthy documents. However, significant effort is required by users or interface developers to associate specific sounds with events or symbols, and further to develop a coherent set these associations into an overall sound scheme for an application.

Other uses of non-speech sound to improve accessibility have been reported in the ICAD literature[7]. These include supporting the navigation of structured documents and the World Wide Web, auditory progress bars, auditory graphs, auditory widgets, auditory

[6] http://www.audiogames.net/
[7] http://www.icad.org/

access to the widgets in a word processor, and use of audio to review data in spreadsheets.

Mobile accessibility

Non-speech sound has been employed in mobility devices; for example, to give longer-range warnings to a blind user of an approaching obstacle than can be obtained using a white cane. A study carried out by Walker and Lindsay (2004), where participants were guided along routes by sound beacons, showed good performance even in the worst case, proving that the non-speech auditory interface could successfully be used for navigation.

The 'K' Sonar (Bay Advanced Technologies, 2006), intended to supplement a white cane, is like a flashlight that sends out a beam of ultrasound instead of light. Using the ultrasound waves reflected from the objects in the beam's path, 'K' Sonar provides complex audible representations of the objects that the user can learn to recognize and build mental maps of their environment.

Despite the success of these applications, problems can arise with the use of audio in such mobility and orientation devices. Sounds that are fed back to the user can be masked or made less clear by ambient sound such as traffic or construction. Although some users might choose to reduce this problem with earphones, many visually impaired users are wary of anything that limits their perception of their environment. This problem can be resolved through the use of a small speaker positioned just a few inches away from one of the ears of the user. The Trekker, developed by Humanware (www.humanware.com), a largely speech-based GPS mobility and orientation system for visually impaired users, uses such a system. This approach keeps the audio output audible and relatively private without masking other environmental sounds. Because of the problems of sound interference, a number of other ultrasound or infrared systems employ haptic feedback rather than sound, using the strength and type of vibrations to provide information about the size and proximity of nearby objects. A promising approach to the sound interference problem is offered by the use of bone-conduction headphones. Rather than going through the eardrum, bone-conduction headphones convert sounds into mechanical vibrations going through the skull straight to the auditory nerve.

5.4 Human Factors Design of an Auditory Interface

Sound is commonly thought of as a way to enhance the display of visual information, but as the preceding sections in this chapter have demonstrated, it also has a range of informational capacities and advantages that make it particularly useful in a variety of interactive settings, and, in some contexts, the best or only choice for conveying information. Thus, the design problem is not just one of evaluating criteria for the use of sound, but is also a question of determining how sound will be used to its best effect.

However, it is important that designers do not utilize auditory displays unless the auditory displays are beneficial and/or necessary. There are many circumstances where auditory displays *could* be utilized, but it is only through the utilization of sound user centered design principles that the practitioner can determine whether an auditory display *should* be utilized.

As with all good user center design, the decision to use sound as a display technique should be based on a comprehensive task and needs analysis and this analysis will inform design specification and the ultimate implementation. This section will provide information about some of the unique issues and challenges practitioners and designers may face when considering auditory interfaces.

5.4.1 Task Analysis—How Does It Differ for an Auditory Interface?

Description of the User

In addition to the usual variables such as gender, age, experience with information technology etc. that developers should consider, the development of auditory displays needs to take specific account of the experience of audio of the target user population for the application as a whole. This in turn breaks down into a number of distinct variables, including:

- Musical expertise—What is the cultural background, level and range of musical expertise of participants? Users with musical training are better in discriminating variations of pitch or temporal features like rhythm.
- Familiarity with auditory displays—irrespective of musical ability, what level of experience will the users have with using auditory displays? (Also see section 5.6.1)
- Hearing abilities—will the user/group of users have a higher likelihood of hearing impairment than other populations (e.g., workers who operate load machinery on a regular basis); do they have limited spatial resolution because of hearing loss in one of their ears; are the users capable of analytical listening?

Work that the user must accomplish

The goal of articulating the specific tasks the user will be performing is twofold. First, it develops an account of both what the user is expected to do in the targeted activity and how that is to be accomplished. Second, it provides detailed descriptions of the information that must be available to the user, both type and number of sources, in order to perform the tasks. Generally, this process entails creating detailed descriptions of the procedures users must follow for accomplishing certain tasks. For auditory displays, however, these procedures are not necessarily visual or psychomotor (e.g., look at the information displayed on the speedometer and adjust the speed accordingly with either the brake or gas pedal). Thus, additional elements of the task must be considered.

1. What Tasks Will the User Perform Using the Auditory Display?

If an auditory display will be part of a larger multimodal interface, which is more often than not the case, the task analysis must account for all facets of the user's task, not just the aural component.

Exploring data with interactive sonification techniques, for example, usually entails more than just interactive listening (Hermann and Hunt, 2005). Many users need iterative access to the data as well as ways to stipulate and revise the manner of auditory

interaction, be this variable parameter mappings or a sonification model. Each of these aspects of the task must be identified and described procedurally. In particular, the use and role of any non-audio interface components, such as a command line or graphical display, common or novel input devices, etc., should be spelled out in the analysis, as well as the kind of cognitive, perceptual, and motor actions the user must perform to define and interact with the sonification. The resulting description of the complete task then serves as the primary basis for specifying the compositional details of the user interface, that is, what the user will actually hear, see, and physically manipulate.

Further, some auditory tasks, particularly those that are driven by external events or data, may be critical components of larger operations that are subject to variable rates of activity or priority. When this is likely to be an issue, it is important to identify potential performance boundary conditions. It may be possible, for instance, for event pacing to exceed the abilities of users to respond to the demands of the task or for the user's attention to be overwhelmed by the priority of other stimuli during periods of high operational activity. Concern with performance limits and task priorities is closely related to other concerns complex user environments raise, but is easily overlooked. By making note of how and when these factors may arise, the task analysis helps to identify where the auditory interface design is likely to require the greatest effort. Indeed, in some situations, auditory displays are used to reduce cognitive load at points in the task when the cognitive load is the highest.

2. What Information Should the Sound Convey?

A detailed account of the information that sound will be used to convey in the auditory interface must also be developed in conjunction with the task analysis. This information analysis specifies what the user must know to achieve his or her application goals, and is especially important in the design of auditory displays because of the somewhat unique challenges auditory information design poses. Unlike other representational techniques that are commonly used in the design of information tasks, particularly those used in the design of visual displays, the mapping of data to, and the perception of meaning in, sound can be subject to a range of individual performance differences (see, in particular, sections 5.1 and 5.3 above).

The information analysis is best organized to correspond to the organization of the task analysis. Descriptions of the information involved at each step in the task should be detailed enough to address the representational and perceptual considerations that will arise at the auditory design stage, many of which have been described in the preceding sections. In particular, the information to be conveyed to user should be characterized in the following ways:

- Its application to elements of the task (one, several, many, etc.) and/or its conceptual purpose
- Its class and organization: Qualitative (nominal, categorical, hierarchical); Spatial (direction, distance); Temporal; Quantitative (nominal, binary, ordinal, integral, ratio, etc.); Range
- Its expected degree of user familiarity (is this information the user will easily recognize or will training be required?)

- A meta-level description of what the information will be used for (e.g., when the information is relevant to other aspects of the task, is redundant or reinforces other displayed information, or has ramifications in some broader context)

An additional facet of the information analysis that is often needed for the auditory information design is an explicit specification of the underlying data (see section 5.2.2 above). This is generally the case for qualitative information and for task-specific subsets of numerical information that are non-linear (e.g., non-contiguous) in one or more ways.

Context and environment of the auditory display

Since certain operational environments have implications for how an auditory display should be organized and presented, the task analysis should also consider the following questions:

First, will the task be performed in a single or shared user space? This question bears directly on how the auditory display is rendered for individual or group use, i.e., through loudspeakers or headphones.

Second, will the user be engaged in more than one task, and if so, what is the role of concurrency? The most important question here is how auditory information in the task being analyzed is likely to interact with the performance of another task the user is expected to perform and vice versa, especially if sound is also used by the other task. Particular attention should be given to the potential for conflicting or ambiguous uses of sound.

Third, will external sound be present in the environment and, if so, what is its informational role? This concern is somewhat like the previous consideration, only at a larger scale. If the auditory task will be embedded in a larger environment in which noise or intentional uses of sound are present, the task analysis should address the potential for masking or other disruptive impacts on the auditory task.

Constraints on Interface Design

After user, task, and environmental considerations, designers of auditory interfaces must consider the practical limits associated with the capacities and costs of current computer and audio technologies. The primary issues designers must confront are processing power and mode of rendering, particularly in middle and lower-tier platforms. There are some environments that will be conducive to a computationally intensive auditory display, e.g., a system that has been engineered to meet the requirements of a virtual immersive environment in a laboratory or a research-and-development facility. However, many handheld devices and portable computers do not have the capacity for a display that requires much computational power, and this will likely be the case for a number of years to come. For instance, many portable devices currently lack the throughput and/or the processing requirements for real-time 3D auditory rendering of multiple sources. Furthermore, widely available solutions for tailoring binaural processing for individual listeners are still several years away. Thus, if the task requires the display device to be

portable, the designer should avoid creating an auditory interface that depends on conveying accurate spatial information to the listener.

Other practical considerations that are unique to auditory interface designs are factors associated with loudspeaker and headphone rendering, personalization, and choice of sound file format. Accurate perception of auditory detail in loudspeaker rendering, for instance, such as location, motion, and high-frequency information, requires the listener to be ideally positioned in relation to the output of the loudspeaker system. Rendering with headphones and/or earbuds may be inappropriate in some operational settings because this can defeat the listener's ability to hear important sounds in the immediate environment. Designers also must consider how users will access and control parameters such as loudness, equalization, program selection, and other features of an auditory interface the user may wish to individualize. A related, ergonomic issue designers must be aware of is the risk of hearing loss associated with repeated exposure to excessively loud sounds. Finally, designers should understand how MIDI (musical instrument digital interface) files work and what the perceptual advantages and functional tradeoffs between uncompressed auditory file formats, such as the WAV and AIFF specifications, and compression formats, such as WMA and MP3, are (see, e.g., Lieder, 2004). If the user of an auditory interface must be able to exchange sounds with collaborators, cross-platform compatibility is a crucial consideration. When this is the case, the designer should insure the interface supports a range of sound file formats that are appropriate for the perceptual requirements of the task. The MP3 format, for example, is adequate for auditory alerts and icons but inferior to the WAV format for many scientific applications.

5.4.2 When to Utilize an Auditory Interface

There are different design motivations for auditory interfaces that result from a detailed task analysis. They can be broadly organized into four functional categories: managing user attention, working with sound directly, using sound in conjunction with other displays, and using sound as the primary display modality. Although each of these have been mentioned and described in previous sections, in this subsection, we will briefly identify specific human factors relevant for each application type. Many of these human factors have been described in some detail in section 5.1, but the usability practitioner may require more information when actually designing and implementing these applications. Thus, additional issues and sources of information are provided here.

Managing User Attention

A variety of auditory materials can be used to manage attention, and the manner in which information is conveyed can be either discrete or continuous, and may or may not involve the manipulation of auditory parameters. For instance, a substantial body human factors research relating the parameters of auditory signals to perceived urgency exists for the design of auditory warnings, which fall under this design category (see Stanton and Edworthy, 1999, for a review). Similarly, changes in the character of continuous or streaming sounds allow listeners to peripherally or preattentively monitor the state of ongoing background processes while they attend to other functions. Whenever designers creates sounds for attentional purposes, the perceptual strengths and weaknesses of the auditory materials must be carefully evaluated in the context of the larger task

environment. In addition, a range of related non-auditory human factors may also need to be considered. These include interruptions, (McFarlane and Latorella, 2002), situation awareness (Jones and Endsley, 2000), time-sharing (Wickens and Hollands, 2000), and stress (Staal, 2004).

Working with Sound Directly

Many activities involve working with sound itself, either as information, as a medium, or both. In this design category, perception and/or manipulation of sound at a meta-level is the focus of the task. For example, a user may need to be able monitor or review a live or recorded audio stream to extract information or otherwise annotate auditory content. Similarly, sound materials, particularly in music, film, and scientific research often must be edited, filtered, or processed in specific ways. Interfaces for sonifying of data, covered at length above, also fall under this heading. Knowledge of human auditory perceptual skills (Bregman, 1990) and processes in auditory cognition (McAdams and Bigand, 1993) are both important prerequisites for the design of any task that involves end-user development or manipulation of auditory materials.

Using Sound in Conjunction with Other Display Modalities

Sound is perhaps most frequently called upon to complement the presentation of information in another modality. Although haptic interfaces are beginning to make use of sound, (see, e.g., McGookin and Brewster, 2006a), more often than not, sound is used in conjunction with a visual display of some sort. Like its function in the real world, sound not only reinforces and occasionally disambiguates the perception of displayed events but also often augments them with additional information. In addition to previously mentioned human factors issues; aesthetics (Leplatre and McGregor, 2004), multisensory processing (Calvert, Spence, and Stein, 2004), and the psychology of music (Deutsch, 1999) could also be relevant for these types of applications.

Using Sound as the Primary Display Modality

In a range of contexts, sound may be the most versatile or the only mode available for representing information. In these auditory applications, a user population with its own, often unique, set of interaction goals and expectations is targeted. Many of the human factors considerations that are relevant for the preceding design categories can also be applicable here, particularly those relevant to working with sound and those relevant to using sound in larger operational contexts.

5.4.3 Specifying the Requirements for an Auditory Interface

An important step in the development process is to turn the products of the contextual task analysis and iterative prototype testing into a coherent specification. Depending on the size and scope of the project, formal methods (e.g., Habrias and Frappier, 2006) or a simple software requirements document can be used for this purpose. However, the value of this exercise and its importance should not be underestimated. The specification is both a blueprint for the interface and a road map for the implementation process.

In whatever form it takes, the specification should detail how the interface will be organized, how it will sound and appear, and how it will behave in response to user input well enough to prototype or implement the auditory task to a point that is sufficient for subsequent development and evaluations. The interface should organize and present the actions and goals enumerated in the task analysis in a manner that is intuitive and easy to understand. Any auditory materials, sound processing specifications, and examples that were developed for the auditory design should be referenced and appended to the specifications document.

In many cases, it will also be effective to sketch the sound equivalent of a visual layout with an auditory prototyping tool. The designer should take care to ensure that the specified behavior of the interface is orderly and predictable, and, because audio can be unintentionally too loud, the interface ideally should cap the amplitude of auditory presentations and should always include a clear method for the user to cancel any action at any point. A number of auditory prototyping tools ranging from simple sound editors to full-blown application development environments are available both commercially and as open source projects that are freely available on the Internet for downloading and personal use. Section 5.2.3 gives programming examples for sonifying data in the open source SuperCollider environment (McCartney, 1996) and also mentions another popular and powerful open source programming environment that can be used for prototyping audio, video, and graphical processing applications is Pure Data (Puckette, 1997). A recent, commercially available tool for developing immersive virtual auditory environments is VibeStudio (VRSonic, 2007).

Last, a good understanding of auditory display technology and the current technology is important for achieving good human factors in auditory tasks. The designer should weigh the advantages of commercial vs. open source audio synthesis and signal processing libraries and should also give attention to the implications of different audio file formats and rendering methods for auditory tasks. Sounds rendered binaurally with non-individualized head-related transfer functions (HRTFs), for instance, are perceived by most listeners to have functional spatial properties but are accurately localized by only a small percentage of listeners. Technical knowledge at this level is integral for developing an effective specification and ensuring that usability concerns remain central during the iterative stages of implementation and formative evaluation that follow.

5.4.4 Design Considerations for Auditory Displays

One of the goals of this chapter's sections on the nature, technology, and current implementations of auditory interfaces is to give the reader a tangible sense of the exciting scope and range of challenges auditory information designs pose, especially with regard to human factors. Auditory design is still more of an art than a science, and it is still very much the case that those who choose to implement auditory interfaces are likely to find they will have to do a bit of trailblazing.

Sections 5.2 and 5.6 give detailed development and design guidelines that the designer and practitioner should find useful when developing sounds for auditory interfaces. Because auditory interfaces utilize sounds in different paradigms than many (indeed most) people are familiar with, the current section is devoted to providing the

reader with a way of thinking about sound from an informational perspective and how the design and production of sounds should be influenced by this different way of thinking.

Thinking about sound as information

In the process of designing auditory materials to convey task-related information, it is important to keep in mind a number of conceptual notions about sound as information. First, sound can be usefully categorized in a number ways, non-speech versus speech, for example, natural versus synthetic, non-musical versus musical, and so on. Listeners generally grasp such distinctions when they are obvious in the context of an auditory task, so this sort of partitioning—in an auditory graphing application, for instance—can be quite useful as a design construct.

Another valuable design perspective on auditory information, introduced in Kramer (1994), is the auditory analogic/symbolic representation continuum mentioned at the beginning of the chapter. Sounds are analogic when they display relationships they represent, and symbolic when they denote what is being represented. Much of the displayed sound information people regularly experience falls somewhere between, and typically combines, these two ideals. The classic Geiger counter example can be understood in this way—the rate of sounded clicks is analogic of the level of radiation, while the clicks themselves are symbolic of radiation events, which are silent in the real world.

The analogic/symbolic distinction can also be usefully conceptualized in terms of semiotics (i.e., the study of symbols and their use or interpretation). When sound is designed to convey information to a listener, the intended result, if successful, is an index, an icon, and/or a symbol (cf. Clark, 1996). Indices work by directing the listener's attention to the information they are intended to signal and icons work by aurally resembling or demonstrating the information they are meant to convey. Both of these types of sounds function in an analogic manner. A remarkable example of an auditory signal that is both indexical and iconic is a version of Ulfvengren's (2003) "slurp," which, when rendered spatially in an airplane cockpit, is intended to draw the pilot's attention to and resemble a low fuel gauge condition. Note how the informational use of this sound in this context also makes it a symbol. Sounds that are symbols work by relying on an associative rule or convention that is known by the listener. Outside of a cockpit, most listeners would not readily associate a slurping sound with their supply of fuel!

A final meta-level aspect of sound-as-information that should be kept in mind as the auditory design process begins is people's experience and familiarity with the meanings of everyday and naturally occurring sounds (Ballas, 1993). Much of the information conveyed by these classes of sounds is not intentionally signalled but is a perceptual by-product of activity in the real world, and is understood as such. Footsteps, the sound of rain, the whine of a jet, the growl of a dog, all have contextual meanings and dimensions that listeners readily comprehend and make sense of on the basis of life-long experience and native listening skills. The inherent ease of this perceptual facility suggests an important range of strategies for auditory designers to explore. Fitch and Kramer (1994) used analogues of natural sounds in a successful patient monitoring application to render concurrent, self-labelling auditory streams of physiological data.

Even more innovative is work by Hermann et al. (2006) in which pathological features in EEG data that are diagnostic of seizures are rendered with rhythms and timbres that are characteristic of human vocalizations.

In contrast, sounds that are unusual or novel for listeners, including many synthetic and edited sounds, as well as music, have an important place in auditory design, primarily for their potential for contextual salience and, in many cases, their lack of identity in other settings. In general, though, unfamiliar uses of sounds require more training for listeners.

Designing the sound

Turning now to practice, once the information to be conveyed by sound has been analyzed, the designer should begin the process of selecting and/or developing appropriate sounds. A useful perspective on the design problem at this point is to think of the auditory content of an interface as a kind of sound ecology (Walker and Kramer, 2004). Ideally, the interface should be compelling, inventive, and coherent—it should tell a kind of story—and the sounds it employs should have a collective identity listeners will have little difficulty recognizing, in much the same way that people effortlessly recognize familiar voices, music, and the characteristic sounds of their daily environments. Good auditory design practice involves critical listening (to both the users of the sounds and the sounds themselves!) and strives foremost to accommodate the aural skills, expectations, and sensibilities listeners ordinarily possess. It is easier than many people might think to create an auditory interface that is unintentionally tiresome or internally inconsistent or that requires extensive training or special listening skills.

Once some initial thought has been given to the organization and character of the listening environment, the first component of the auditory design process is to work out how sound will be used to convey the task-related information that is identified in the task analysis. Often, it is also useful to begin developing candidate sounds for the interface at this time, because this can help to crystallize ideas about the design; however, this may not always be possible. The mapping from information to sound should, in many cases, be relatively straightforward, but in other cases, for instance with complex data relations, it will generally be necessary to experiment with a number of ideas. Here are several examples.

- Event onsets intuitively map to sound onsets.
- Level of priority or urgency can be represented systematically with a variety of parameters including rhythm, tempo, pitch, and harmonic complexity (e.g., Guillaume, Pellieux, Chastres, and Drake, 2003).
- Drawing attention to, or indexing, a specific location in space—a form of deixis (Ballas, 1994)—can be accomplished with three-dimensional audio rendering techniques.
- Emotional context can be conveyed with music or musical idioms.
- Distinct subclasses of information can be mapped to different timbres; ranges can be mapped to linearly varying parameters.

- Periodicity can be mapped to rhythm.

Many more examples could be given. Often, there will be more than one dimension to convey about a particular piece of information and in such instances auditory parameters are frequently combined. An auditory alert, for example, can pack onset, event identity, location, level(s) of urgency, duration, and confirmation of response into a single instance of sound (Brock, Ballas, Stroup, and McClimens, 2004).

Producing the sound

As mappings and candidate sounds for the interface are developed another factor the auditory designer must address is how the final sounds will be produced, processed, rendered. Although an introduction to the technology of sound production is given above in section 5.2, the emphasis there is primarily on computational techniques for the synthesis of sound. Other means of sound production include live sources and playback of recorded and/or edited material. In addition, many auditory applications require sounds to be localized for the listener, usually with binaural filtering or some form of loudspeaker panning. And some tasks allow or require the user to control, manipulate, assign, or choose a portion or all of its auditory content. Consequently, vetting an auditory design to ensure that its display implementation will function as intended can range from assembling a fixed set of audio files for a modest desktop application to specifying a set of audio sources and processing requirements that can have substantial implications for an application's supporting computational architecture. In the latter situation, one would reasonably expect to be part of a collaborative project involving a number of specialists and possibly other designers.

Choosing between one construct and another in rich application domains and knowing what is necessary or most likely meet the user's needs is not always a matter of just knowing or going to the literature. All of these advanced considerations, however—the production, filtering, augmentation, timing, and mixing of various types and sources of sound—are properly part of the auditory design and should be identified as early as possible in a complex design project because of their implications for the subsequent implementation and evaluation phases of the auditory interface.

5.4.5 Iterative Evaluation

The final and indispensable component of the auditory design process is formative evaluation via user testing (Hix and Hartson, 1993). Targeted listening studies with candidate sounds, contextual mock-ups, or prototypes of the auditory interface, designed to demonstrate or refute the efficacy of the design or its constituent parts, should be carried out to inform and refine iterative design activities. For more on evaluation, see section 5.5 below.

5.5 Techniques for Testing the Interface

As with every interactive system the evaluation of auditory displays should ensure a high degree of usability. For auditory displays, finding methods to evaluate usability is not a trivial task. The subsequent sections will highlight some of the specific approaches and issues relevant to the evaluation of auditory displays.

5.5.1 Issues specific to the evaluation of auditory displays

Early prototyping: There are few generally available tools for the early prototyping of concepts in auditory displays, but the desirability of obtaining early feedback on auditory interface designs is, if anything, even more important than in prototyping visual displays because many users are relatively unfamiliar with the use of audio. Wizard of Oz techniques (Dix, Finlay, Abowd, and Beale, 2004) and the use of libraries of sounds available on the Internet (FindSounds, 2006) can provide the basis of ways around this dilemma. The quality and amplitude of the sounds employed in prototypes must be close to those anticipated in the final system in order to draw conclusions about how well they are likely to work in context. For instance, research by Ballas (1993) shows that the way an individual sound is interpreted is affected by the sounds heard before and after it, so accurate simulation of the pace of the interaction is also important. One early stage technique that can be used is vocalize what the interface is expected to sound like; for example, Hermann et al. (2006) used such a technique to develop sonifications of EEG data. By making it possible for people to reproduce the sonifications, users were able to more easily discuss what they heard and these discussions facilitated the iterative testing process.

Context of use: Evaluation of the display in the environment where the system will be used and in the context of users performing their normal tasks is particularly important for auditory displays. The primary factors associated with context are privacy and ambient noise. For instance, although the task analysis (described in section 5.4.1) may have determined that privacy is necessary and thus, headphones are the best delivery method for the sounds, an evaluation of the display in the environment may determine that wearing headphones interferes with the users' task. Conversely, if privacy is not an issue and speakers are being used for the auditory display, an evaluation of the display in the context where it will be used could determine the appropriate placement and power of speakers.

Cognitive load: If an auditory display is being used to reduce cognitive load, the evaluation process should confirm that the load reduction occurs. One way to measure cognitive load is Hart and Staveland's NASA Task Load Index (TLX: 1988). If this measure is not sensitive enough for the tasks associated with the auditory display, it may be better to use accuracy and or time performance measures as indirect measures of whether the display has decreased cognitive load.

Choice of participants: When conducting evaluations of auditory display, as with all evaluations of all types of displays, the characteristics of the participants should match those of the intended users as closely as possible. Some of the obvious variables that should be considered are gender, age, and experience with information technology. Furthermore, evaluators need to also match participants on the specific variables associated with audition (listed in 5.4.1).

When conducting evaluations, there are also dangers of making false assumptions concerning the applicability of evaluation data across different user types. For example, it is not unusual in the development of systems for visually impaired users for sighted users, who have had their view of the display obscured in some way, to be involved in the evaluation. However, in a study involving judgments concerning the realism of sounds and sound mappings, Petrie and Morley (1998) concluded that the findings from sighted

participants imagining themselves to be blind could not be used as a substitute for data from participants who were actually blind.

Finally, evaluators need to be particularly diligent about determining whether participants have any hearing loss. Obviously, hearing losses are likely to impact participants' interactions with the system and thus should be controlled.

Data capture: As might be expected, problems can arise with the use of think-aloud protocols for capturing the results of formative evaluations of auditory displays. Participants are likely to experience problems when asked to articulate their thoughts while at the same time trying to listen to the next audio response from the interface. This is not to rule out the use of think-aloud protocols altogether. For example, Walker describes in relation to the evaluation of Mobile Audio Designs (MAD) Monkey, an audio augmented reality designer's tool, there may be situations where the audio output is intermittent and allows sufficient time for evaluators to articulate their thoughts in between audio output from the system (Walker and Stamper, 2005). Another alternative would be to use what is commonly known as retrospective think-aloud protocol, in which participants describe the thoughts they had when using the system to the evaluator whilst reviewing recordings of the evaluation session.

Learning effects: Improvement in performance over time is likely to be important in most systems, but there is currently little known about learning effects observed in users of auditory displays, other than the fact that they are present and need to be accounted for: In experiments conducted by several of the authors, significant learning effects have frequently been seen early in the use of auditory displays as users transition often from never having used a computer-based auditory display before to gaining some familiarity in reacting to the display. For instance, a study by Walker and Lindsay (2004) of a wearable system for audio-based navigation concluded "practice has a major effect on performance, which is not surprising, given that none of the participants had experienced an auditory way-finding system before. Thus it is critical to examine performance longitudinally when evaluating auditory display designs."

Heuristic evaluations: It is one of the great advantages of sound that the auditory cues employed can be designed to be background noises; hence auditory displays are often used as ambient displays. Mankoff and her colleagues (2003) developed heuristics for revealing usability issues in such ambient displays. Heuristic evaluation of user interfaces is a popular method, because it comes at very low costs. For example, Nielson found that a panel of 3 to 5 novice evaluators could find 40 - 60% of known issues when applying heuristic evaluation (Nielsen and Molich, 1990). However, doubts have been expressed about the results of some studies investigating its effectiveness and some usability professionals argue that Heuristic evaluation is a poor predictor of actual user experience, see for example http://www.usabilitynews.com/news/article2477.asp.

5.5.2 An example of a cross modal collaborative display: Towers of Hanoi

One common concept in accessibility is that given a collaborative situation between sighted and visually impaired users is that the difference in interaction devices can cause problems with the interaction. Winberg and Bowers (2004) developed a Towers of Hanoi game with both a graphical and audio interface to investigate collaborative work between sighted and non-sighted work. To eliminate any problems associated with having

different devices for sighted and blind users, both interactions were mouse based, employing a focus feature to enable the mouse to track the cursor. The sighted player worked with a screen, the blind one had headphones. In order to encourage collaboration, there was only one cursor for the two mice.

Testing set-up: In order to keep things equal, neither player had access to the other's interface. The sighted player had a screen and could see the blind participant but not his or her mouse movement; both could hear each other, as this was necessary for collaborative work. Each player was trained independently and had no knowledge of the other's interface.

Evaluation: The entire interaction between the two players was video taped and the game window was also recorded. The video enabled Winberg and Bowers (2004) to study the entire interaction and the screen capture allowed them to see the state of the game at all times. The players played three games: respectively with 3, 4 and 5 disks.

Analysis: In this experiment in cross modal collaboration, Winberg and Bowers studied the following aspects of the interaction: turn taking, listening while moving, monitoring the other's move, turn taking problems and repair, re-orientation and re-establishing sense, engagement, memory and talk and disengagement. The major method used to evaluate the interaction was Conversation Analysis (tenHave, 1999). The transcription and subsequent study of the conversation paired with the players actions gave an in depth qualitative analysis of problems in the interaction. Any problems with the interfaces became apparent from stumblings and confusion in the conversation. Actions were also timed, and this helped to pinpoint problems with the direct manipulation in the auditory interface.

Conclusions: The system developed and evaluated by Winberg and Bowers (2004) enabled the examination of some basic issues concerning the cooperation between people of different physical abilities supported by interfaces in different modalities. They concluded that sonic interfaces could be designed to enable blind participants to collaborate on the shared game: in their evaluation, all pairs completed all games. The auditory interface enabled blind players to smoothly interleave their talk and interactions with the interface. The principle of continuous presentation of interface elements employed in the game allowed blind players to monitor the state of the game in response to moves as they were made. This enabled to blind player to participate fully in the working division of labor. Both blind and sighted collaborators therefore had resources to monitor each other's conduct and help each other out if required. However, problems were seen when the blind player stopped manipulating the display and listening to the consequent changes. In these situations the state of the game became unclear to the blind player and difficulties were experienced in re-establishing their understanding of the current state of the game.

Important findings resulting from the study by Winberg and Bowers (2004) can be summarized as follows:

1) The manipulability of an assistive interface is critical, not only for the purpose of completing tasks, but also to enable a cross modal understanding of the system state to be established. This understanding can become compromised if the linkage between gesture, sound and system state becomes unreliable.

2) When deciding whether to implement functionality in sound, the availability of other channels of communication and the appropriateness of audio for representing the function should be kept in mind. For example there could be situations where the sonification of an interface artifact may simply take too long and may be better replaced by talk between collaborators.

3) It is not enough to design an assistive auditory interface so that it facilitates the development of the same mental model as the interface used by sighted individuals. Additionally, it is essential to examine how the assistive interface will be used and how this usage is integrated with the various things that participants do such as "designing gestures, monitoring each other, establishing the state of things and one's orientation in it, reasoning and describing" (Winberg and Bowers, 2004). In order to do this effectively, it becomes essential to study how people use assistive interfaces in collaborative situations.

5.6 Design Guidelines

Although the potential of audio as an interaction modality in HCI is high and many applications have shown this (e.g., Brewster, 2002), the efficient design of audio remains something of a mysterious process and guidance is often scarce. Hence, the remainder of this section is an attempt to describe the existing guidelines, principles and design theories.

5.6.1 Analysis and requirement specification

In auditory design some aspects of the requirement specifications demand special attention. As mentioned in 5.4.1, it is important to have a clear understanding of the listening background and abilities of the users. In addition to the issues listed in 5.4.1, designers should considered the users' openness to an alternative display modality. Audio as part of human-technology interaction is a comparatively new field and thus, users have little experience with using it. This also means that designers may encounter scepticism and prejudice against using audio; not only from users, but from all stakeholders in the design process. Although there might be strong arguments for using audio in a specific application, a client might still request a visual solution because he or she cannot imagine an auditory solution.

A valuable concept for auditory information design in this early phase was proposed by Barrass and is called TaDa-Analysis (Barrass, 1997). It is a method for describing the task and the data to be represented in a formal way including a story about the usage and properties that are decisive for auditory cues like attention levels or data types. Barrass used these TaDa descriptions as a starting point for the selection of sounds and then employed these descriptions to match them with sounds stored in a database (EarBender). The EarBender database contains a large number of sounds tagged with semantic and other properties that can be matched with the requirements from a TaDa analysis (Barrass, 1997). Barrass also proposes the creation of auditory design principles based on principles for generic information design like directness or the level of organization. He links these with the properties of auditory perception to create auditory design principles (Barrass, 1997).

The TaDa technique has been used by a number of auditory designers. Notably, at the "Science by Ear" workshop that took place at the Institute of Electronic Music (IEM) in Graz, in 2006, the technique was used to formalize requirements for a number of case studies for which multi-disciplinary teams were formed to design data sonifications. The case studies included data drawn from Particle Physics, electrical power systems, EEG data, global social data and rainfall data. The TaDa technique proved helpful in providing a standard format for representing the information requirements of each sonification to be designed for use by the multi-disciplinary teams. Examples of the case studies employed at the workshop can be found in the papers by De Campo et al. presented at the International Conference on Auditory Displays (ICAD: 2007).

Of course, other methods and guidelines can and should be applied at the requirement specification stage of the design. Task analysis, user scenarios, personae and other concepts have been successfully applied to visual design and do not involve the need to specify any interaction modality. Examples where these concepts have been applied in auditory designs are the use of rich user scenarios in the design of an auditory web browser (Pirhonen, Murphy, McAllister, and Yu, 2006) and the first stage in the design methodology proposed by Mitsopoulos (2000). Both approaches are elaborated in the next section.

5.6.2 Concept design

Concept design is when high-level design decisions are made while leaving most details still unspecified. This phase links the design problem with concepts of auditory displays. The first and foremost task in this phase is to decide which parts of the user interface audio will be used for and which auditory concepts match the requirements and constraints defined in the requirements phase.

Brewster addressed this issue in a bottom-up approach: find errors in individual parts of an existing interface and try to fix them by the addition of sound. He adopted the event and status analysis and extended it to be applicable to different interaction modalities (i.e. to accommodate audio). This was an engineering approach to reveal information hidden in an interface that could cause errors. Brewster suggested using sound to make this information accessible and linked the output of the analysis to his guidelines for the creation of earcons (Brewster, 1994).

As mentioned in the previous section, Pirhonen et al. (2006) proposed a design method that linked user tasks and auditory cues through the use of rich use case scenarios. The use case was developed with a virtual persona that represented the target group and told the story of how this persona carried out a specific task. The story was enriched with as much detail about the environment and the background of the user as was possible to create a compelling scenario; the authors proposed that “the use scenario should have qualities that enable the interpreter (to) identify him/herself with the character” (Pirhonen et al., 2006, p. 136). Then a panel of 4-5 designers went through this scenario and tried to produce descriptions of sounds to support the task. After creating the sounds as suggested by the designers, another panel was organized and went through the use case scenario that was enriched by the initial sound designs. This procedure was iterated until a working design was found. The method stressed the importance of linking

the user tasks with the design, but also relied heavily on the availability of expert designers for a panel, their experience and ideas.

Another tool for concept design is the utilization of design patterns (Frauenberger, Holdrich and de Campo, 2004). However, there are not yet enough successful implementations of or methodological frameworks for auditory displays for the incorporation of design patterns in the auditory design process. Nevertheless, this tool most likely will prove beneficial in the future.

Another, more theoretical approach has been proposed by Mitsopoulos, founding his methodology on a framework for dialogue design (Mitsopoulos, 2000). Mitsopoulos' methodology consists of three levels: 1) the conceptional level in which the "content" of the interface is specified in terms of semantic entities, 2) the structural level in which sounds are structured over time and 3) the implementation level in which the physical features of the sound are determined. Mitsopoulos proposes guidelines for each of these levels that are derived from theories of auditory perception and attention (e.g. Arons, 1992; Bregman, 1990). By applying these theories he intended to narrow the design space by eliminating designs that would violate psychological principles. Notably, he argued for two fundamental modes of presentation of information by audio: fast presentation, i.e. "at a glance" and the interactive presentation for more detailed user interaction. Each representation is defined in all three levels. Although Mitsopoulos' methodology and guidelines are properly founded in theory, it is important to note that the approach requires a steep learning curve.

In general, decisions in the concept design phase are crucial for successful auditory design, but there is little guidance available that may help novice designers. It is important to note that most flaws in auditory designs are founded in design decisions that occur during the conceptional phase as they tend to be overly influenced by visual thinking. Good auditory design gives prominence to the characteristics and strengths of audio and adopts visual concepts only if there is evidence that they work in the auditory domain.

5.6.3 Detail design

Many specifications that are the results of the prior design stage describe the sounds vaguely, or only some of its properties. In the detailed design stage, these specifications are mapped onto physical properties of sound.

The book produced by Kramer in 1994 is often seen as a landmark publication in Auditory Display design. It reported the proceedings of the first meeting of the ICAD in 1992, including a CD of audio examples illustrating many of the Psychological phenomena, techniques and applications discussed. Several chapters in the book present principles for use in representing information in audio. The book presents a number of methods for associating perceptual issues in auditory display with techniques for their practical implementation. Kramer introduced some of the fundamental sonification techniques such as the direct representation of data in sound (or audification), as well as a number of approaches to mapping data variables into a range of sound parameters such as pitch, loudness, timbre, tempo, etc. The book also provided an overview of many other relevant issues in auditory display design. Examples of these include:

- As would be expected, concurrency is an issue in auditory display design. Clearly there is a limit to how much auditory information human beings can perceive and process concurrently, nevertheless concurrency is potentially a powerful tool in auditory display design, as evidenced by the ease with which even untrained musicians can detect one instrument playing out of tune in a whole orchestra of players.
- Metaphor, as in the rest of user interface design, can be an effective mechanism for developing and supporting the user's mental model of the system. See the use of the group conversation metaphor to support multi-tasking described in the Clique case study (section 5.7.2) as a particularly effective example of this.

Much of the book focuses on applications involving the design of sonifications of complex data, i.e. applications representing either raw data or information to be presented in sound. Additionally, there is also a good deal of valuable guidance in the book for those involved in the design of more symbolic auditory interface elements.

Gaver provided a clear user-interface focus in the same book (Gaver, 1994). He presented techniques to create auditory icons for user interfaces in computing systems. As mentioned previously, auditory icons are based on our everyday hearing experience, thus, familiarity and inherent meaning-making make them highly efficient auditory cues. Hence, when creating auditory icons, they are not described in the usual dimensions of sound like pitch or timbre, but according to properties of the real-world object that causes the sound. With regard to detail design, auditory icons can be parameterized by dimensions like material, size or force and when synthesizing auditory icons, designers seek to use algorithms that allow them to influence these instead of the physical properties of the sound directly (e.g. pitch, loudness etc.). Gaver provides a wide range of such algorithms for impact sounds, breaking, bouncing and spilling effects, scraping and other machine sounds.

Blattner et al. developed guidelines for constructing earcons based on visual icons (Blattner, Sumikawa, and Greenberg, 1989). In their terminology, representational earcons are similar to auditory icons and are built on metaphors and inherent meaning. For abstract earcons, they used musical motifs (a brief succession of ideally not more than 4 tones) as a starting point and defined rhythm and pitch as the fixed parameters. Timbre, register and dynamics were the variable parameters of the motif. By systematically altering the fixed parameters designers could create distinctive earcons while altering the variable parameters would produce earcons with perceivable similarity and may be used to create related families of earcons. In their guidelines, Blattner et al. suggest choosing the tones according to the cultural background of the target group, e.g. Western tonal music and they elaborate on exploiting hierarchical structures in earcon families for better learn-ability. Such compound earcons can be created through combination, transformation and inheritance of one-element earcons.

In his work on guidelines for creating earcons, Brewster refined the guidelines mentioned above and provided more specific guidance regarding rhythm, timbre, pitch etc. (Brewster, 1994). Key guidelines given by (Brewster, 1994) are as follows:

- Use musical instrument timbres to differentiate between earcons or groups of earcons as people can recognize and differentiate between timbres relatively easily.
- Do not use pitch or register on their own to differentiate between earcons when users need to make absolute judgements concerning what the earcon is representing.
- If register must be used on its own then there should be a difference of 2 or 3 octaves between earcons.
- If pitch is used it should not be lower than 125 Hz and not higher than 5 kHz to avoid the masking of the earcon by other sounds and be easily within the hearing range of most users.
- If using rhythm to distinguish between earcons, make the rhythms as different from each other as possible by putting different numbers of notes in each earcon.
- Intensity (loudness) should not be used to distinguish between earcons as many users find this annoying.
- Keep earcons short in order not to slow down the user's interaction with the system.
- Two earcons may be played at the same time to speed up the interaction.

He also investigated the concurrent use of earcons and McGookin and Brewster summarized some of the issues with using concurrent audio presentations in auditory displays (McGookin and Brewster, 2006b). Lumsden and her colleagues provided guidelines for a more specific scenario, i.e., the enhancement of graphical user interface widgets such as buttons by earcons (Lumsden, Brewster, Crease, and Gray, 2002). Although a detailed description of the design guidelines of these additional considerations for earcons is beyond the scope of this chapter, the studies by Brewster with McGookin and Lumsden are a good resource for the earcon designer.

For more on auditory information and interface design, a number of excellent resources can easily be found on the World Wide Web including some mentioned in this section. De Campo (2007) presents a useful design space map for data sonification and references numerous examples that are available at the SonEnvir project website (http://sonenvir.at/). A new online edition of The Handbook for Acoustic Ecology (Truax, 1999) provides an invaluable glossary for acoustic concepts and terminology, as well as hyperlinks to relevant sound examples. Additionally, a wealth of research papers and other resources for auditory design as well as an active online design community can be found at the ICAD website.

5.7 Case Studies

5.7.1 Clique - enhanced screen-reading

Clique[8], developed by Peter Parente is an effort to increase the accessibility of graphical user interfaces for the visually impaired. It is included here as a case study particularly because of the design approach taken. The approach, which was in part informed by target users (blind screen reader users) seems to us to capture particularly well a number of key elements that should inform auditory display design.

The state-of-the-art technologies available for blind computer users are screen-readers and soft Braille[9]displays limited to 1 or 2 lines of display. Screen readers are preferable in many situations as they are considerably less expensive, can run using standard sound cards and many blind users are unfamiliar with Braille. In spite of these strong reasons for their preference, Screen readers are very serial in the way they present information and therefore can introduce usability problems and put their user group at a disadvantage when they are used with GUIs that exploit spatial layout. The motivation for developing Clique was to improve this situation by utilizing a more sophisticated and audio focused design that would exploit the capabilities of human hearing to a better degree and hence improve accessibility. To evaluate how well this was accomplished, as part of the development process, a formal user study was conducted using a sighted human mediator between the computer and users to investigate what would be the most natural form of interaction. This provided a variant on the well-known Wizard of Oz technique where, instead of simulating the functionality of a still to be built visual display, the human mediator simulated a still to be built auditory interface.

The key requirements, obtained by discussions and trials with visually impaired users, were identified as follows:

- The system is based around user tasks rather than visual representations of information or the layout of GUI objects. The system departs from previous screen reader approaches that seek to mimic visual interfaces, focusing instead on how best to assist users in completing their tasks. A multi-channel environment is employed in which application tasks are mapped to virtual assistants.
- Both speech and nonspeech are employed. This approach seeks to reduce the problem encountered with many screen readers where the predominant use of a single speech stream leads to a mismatch in bandwidth between the auditory and visual domains. To further assist in closing this gap in bandwidth, auditory information is sometimes presented concurrently in order to maximise the capabilities of human hearing.
- To be usable with a wide range of applications, Clique employs existing software interfaces and feature semi-automatic generation of auditory representations.
- Finally, Clique has not been developed as a niche product for visually impaired users, but has a broader audience in mind who might benefit from efficient non-visual interaction in situations where visual interaction is difficult or impossible.

[8] http://www.cs.unc.edu/~parente/clique/
[9] Braille is a tactile method that is widely used by blind people to read and write.

Clique is designed to interpret the various objects, relationships and metaphors on the visual display, presenting their meanings in forms appropriate to audio. This method of interaction removes from the user the need to control applications via their native GUIs.

This objective is achieved through the use of scripts that link auditory representations for tasks with information corresponding to those tasks exposed by the platform accessibility API (e.g. Microsoft Active Accessibility). "For instance, the script for Microsoft Outlook Express contains a task definition for browsing email. This definition contains auditory views for browsing a hierarchy of mailboxes, a list of message headers, and the text of an email body. The script associates these views with adapters for the GUI tree view, multi-column list, and text area widgets showing the relevant information on the screen. The auditory views draw information from these adapters and present it to the user. This model-view separation allows Clique to create a consistent auditory experience across applications by reusing the auditory views to represent equivalent tasks that have differing visual representations. [10]" The development approach taken in Clique allows for the definition of interaction patterns that might recur in other applications and so supports re-usability of designs—a highly desirable feature.

User tasks were identified in four target applications—Outlook Express (email), Firefox (web browser), WinZip (archive utility), and Day by Day Processional (calendar)—Parente based his design rationale for the auditory representation of these tasks on reviews of relevant literature and consulting with both visually impaired and sighted users. Auditory icons are used to indicate expected interactions (e.g. list of items, editable text) and common states (e.g. misspelled words). Earcons are used to represent related messages about changes in application state (e.g. task starting, task ending, task interruptions). Ambient sounds are used to represent the user's current working context (i.e. active application, active task in that application). Speech is used to give details about the current task and offer further explanations of the non-speech sounds on demand. The specific design of the sounds was informed by common guidelines and principles such as Blattner et al.'s (1989) guidelines for the presentation and audio parameters of earcons, Bregman's (1990) investigations of the design of complex auditory scenes, Brewster (1994) guidelines for the design of earcons, Gaver's (1994) guidance on the design and use of auditory icons and Mynatt's (1995) recommendations organization and presentation of auditory objects in a screen reader. An example of how these guidelines is seen in Clique's adoption of a group conversation metaphor to facilitate user interaction with multiple concurrent tasks within an audio interface. Multiple virtual assistants are "placed around the user in a virtual sound space. Each assistant is assigned a specific role in the conversation, speaks with a unique voice, plays audio icons to indicate important events, answers user questions, and carries out user commands. Natural constructs of conversation are used during the interaction including references, grounding, pacing, turn-taking, interruptions, and simultaneous speaking."

Clique is implemented in the Python interpreter language[11] that can interface with C libraries and so is able to communicate with all major accessibility interfaces of operating systems and integrate a powerful sound synthesis library. The current prototype

[10] http://www.mindtrove.info/oss/clique.html
[11] http://www.python.org

interfaces with the Microsoft® accessibility interface[12], uses the Microsoft® Speech API[13] for synthesizing speech, and the FMOD[14] library for managing concurrent sound streams and spatialization.

The evaluation of Clique involved testing against the heuristics for ambient displays developed by Mankoff et al. (2003). This is a particularly inexpensive evaluation that allows for the alteration of some usability problems before going into user testing. Subsequently, Clique was evaluated in two summative user studies; one testing the simultaneous speech streams, semi-modal search features, memory aids, and task-based structures with users of a conventional screen reader and the other focused on how efficiently sighted users can use desktop workstations.

5.7.2 Nomadic Radio - wearable personal assistant

Nomadic Radio, developed by Sawhney and Schmandt (2003) at the MIT Media Labs, is a wearable device that relays asynchronous communication to the user, such as voicemail, email, news alerts, and agenda information. With a computer, Personal Digital Assistant (PDA) or mobile phone, users are often required to stop what they are doing in order to check new messages. Continuous interruptions can delay or halt users' work in progress. These interruptions should be strategic and selective particularly due to the risk that the task a user was working on when interrupted may not be picked up again for some time. Nomadic Radio seeks to solve this problem through a hands-free audio interface that keeps track of recent activities and the current environment to determine the urgency of the message and thus how to alert the user.

The design of Nomadic Radio was based on providing quick and non-disruptive access to various sources of messages and personal information. An auditory display is immediately attractive as this display can easily be absorbed in the background without interfering with the current task and can also be easily ignored if there are more important things to which to attend. Nomadic Radio's primary input modality is speech, making it is easy for users to command the system without interrupting their task. The system understands a small set of commands designed to be easily remembered and distinguished: *"Go to my {email | news | calendar | voice-mail}", "Move {forward | back},"* etc. The primary output modality is also speech. Messages and summaries are read out to the user on command.

Nomadic Radio also makes use of non-speech sound in the form of audio cues and ambient sounds. Audio cues provide general feedback and indicate the priority and category of the message. The audio cues describe the type of message and the priority is determined by content filtering. Another key aspect to Nomadic Radio is the spatialization of the audio. Messages can be played simultaneously and ones of importance move to the foreground. The audio cues attract the users' attention in such a manner that they can decide whether to focus on it. Nomadic Radio also uses ambient sound to continuously report the state of the system. The sound of water is used to indicate activity: a gentle flow indicates low activity, a splash is the arrival of a short message, and larger messages cause the flow to become more agitated. The changes in

[12] http://msdn.microsoft.com/at/
[13] http://www.microsoft.com/speech/default.mspx
[14] http://www.fmod.org/

pattern of the water sound can prepare the user for upcoming messages. If the system determines that the user does not want to be interrupted, it provides an easy non-disruptive way for the user to still be aware of the content coming in. This is a significant contribution to auditory interfaces as it addresses the annoyance factor in a way that can be controlled and mitigated by the user. The soft sounds prepare the user for messages and also allow them to turn off the system if it is inconvenient.

Nomadic Radio was developed as an advanced prototype. As such, its creators, Sawhney and Schmandt, have primarily used it. They performed two short evaluations with 3 users. The users were asked to use the system for short periods of times over a period of several days. The feedback that pertained to the non-speech audio uncovered several things:

- The sound notifications such as introduction to messages were successful in conveying information to users and in fact could be attended to while performing other tasks and even holding conversations. Here, we can see a successful support of multitasking with the use of ambient sound.
- Volume control was very important to users: they liked turning off the sound for meetings, and being able to turn up and down the volume depending on surrounding sounds and their desire for privacy. Here, we can see that flexibility is important to users.
- One user strongly preferred longer and gradual notifications to short ones. Here, we see how more abrupt sounds are more disruptive, which should be reserved for high priority messages.
- All the users preferred for the system to make some noise all the time to reassure them that the system is still operating. Here, we see the importance of ambient noise, such as one would hear during lulls in a telephone conversation, to reassure the user. This can also be done by providing an easy mechanism to probe the system.

Nomadic Radio uses Java clients to communicate wirelessly over a Local Area Network (LAN) with remote servers written in C and Perl. The information being presented is handled by *PhoneShell* (Schmandt, 1993), a system that allows remote access to desktop information. The audio is rendered with the RSX 3D audio API[15] that uses Head-Related Transfer Functions (HRTF) to place the streams. The speech input and output is handled by the AT\&T Watson Speech API (Goffin et al., 2005). The wearable device is the *SoundBeam Neckset*, a prototype from Nortel, which rests around the neck. Speakers on each shoulder provide the audio and allow for spatialization of the audio. The directional microphone is mounted on a solid tongue that curves down from the right shoulder to rest on the chest. The computer, worn on the hip, is a Toshiba *Libretto* PC--a mini-computer about the size of a VHS tape -- running Windows® 95/NT.

5.8 Future Trends

The material presented in this chapter has illustrated that *auditory interfaces* are versatile and efficient means of establishing a communication between a computer system and the

[15] http://developer.intel.com/ial/rsx/index.htm

user. Listening is perhaps the communication channel with the highest bandwidth after visual perception, and certainly a channel whose characteristics are so different from visual perception that particularly the combination of visual and auditory displays covers a very wide range of interface possibilities. Whereas vision is a focused sense (we only see where we look at), sound surrounds the user; while vision stops at surfaces, sound allows us to discover 'inner' worlds, beyond the visible limits; whereas vision is persistent, sound is intrinsically aligned to changes over time, to dynamic interaction, and our auditory skills are particularly good at discerning dynamic properties in sound.

Looking at auditory interfaces from a more distant view, we see that two directions are possible: to use sound as a display or as an input modality. This chapter focused particularly on the display mode. However, the input mode can be equally compelling. An interface could employ speech (symbolic) and non-speech (analogic) as auditory inputs. The latter has potential for development hand in hand with the development of auditory interaction systems using sonification. For instance, systems that allow a user to tap a rhythm (picked up via a microphone) for selecting the tempo of music tune would fall under the analogic category. Query-by humming systems would be an example for auditory non-symbolic input where the auditory input would be used to select specific music. Sonification is currently laying out new avenues for the use of non-symbolic auditory inputs. For instance, Hermann et al. (2006) presented vocal sonification of EEG, inspired by the excellent human capabilities of discerning and memorizing structure in vocal sounds. Since humans are furthermore able to mimic vocal patterns, they can use their own vocal tract to actively reference certain patterns in the sonification. This process may simplify the communication of data patterns.

Auditory interfaces, both as input and output, each both as analogic and symbolic interface, are likely to gain relevance in future user interfaces to come for several reasons: First, because the technological development is just starting to enable these interactions at a sufficient level of sophistication, and second, because there are many situations where the visual sense is not available or otherwise used, and finally, because we would simply be wasting an excellent and highly-developed communication channel if sound is neglected in the user interface.

Additional future trends that may develop in auditory displays are: a) *interactive* sonification—a better closure of interaction loops by interactive sonification, and b) an amalgamation of auditory display with other modalities such as visual display and tactile/haptic interfaces that would result in truly *multi-modal interfaces*.

Interactive Sonification (Hermann and Hunt, 2005) bears the potential to create intuitive control of systems at a level beyond a rational (logic) analysis of steps, more as an intuitive, creative, and synergetic approach to solve problems. For instance, in data analysis via interactive sonification, discovering patterns would turn from a step-by-step analysis into a continuous movement through the data space or sonification space. The user would integrate any locally gained insight regarding the structure of the data into his or her exploratory activity in a continuous way, without disrupting the activity and experience. Such a continuous, interruption-free mode may better create the experience of *flow*, the dissolving of a user in the activity, which in turn may give a better access to the user's often covered creative potential. Multi-modal interfaces, on the other hand, will allow the designer to combine the strengths of different modalities, so that the

communication of data is simplified and achieves a better matching to the user's perception capabilities. For instance, if the user's visual focus is already highly used, multi-modal display engines could automatically select auditory components, or even just emphasize them against the visual counterpart, so that the communication between the user and a computer system is optimized.

Another trend in auditory interfaces that is gaining momentum is the concept of intelligent or adaptive auditory environments and displays. Advances in rendering, signal processing, user modeling, machine listening (Wang and Brown, 2006), and non-invasive user monitoring technologies mean that in the relatively near future, many interactive devices and environments will transparently adapt the audio component of their information displays to match the needs of users, much like people rely on each other's listening skills in social settings to coordinate the aural dimension of conversation and other shared forms of sound information.

Three areas in which adaptive sound technology is already being explored are mobile telephony, pervasive computing, and social robotics. Mobile phones have arguably become the most common auditory interface people encounter in their day-to-day lives. To compensate for noise in dynamic environments, new wireless headsets are already being marketed that adaptively alter a mobile phone's outgoing and incoming audio signals to improve speech communications (see, e.g., Mossberg, 2006). As mobile phones move beyond telephony in to areas as diverse as internet access, personal entertainment, content creation, and interaction with so-called smart and pervasive computing environments, exciting new opportunities for intelligent auditory presentation behaviors are arising. In recent pervasive computing research, for instance, users intuitively navigated their way to undisclosed outdoor locations using a context-dependent, directionally adaptive auditory display (Etter and Specht, 2005). The underlying system uses global positioning data and a geographical information system to infer the mobile user's geographical context. Navigation cues are then rendered by adaptively panning and filtering music selected by the user to correspond with his or her direction of travel. Intelligent, adaptive auditory displays are also expected to be an important technology for social and service robots. Recent human-robot interaction work by Martinson and Brock (2007) explores several strategies for adaptively improving a robot's presentation of auditory information for users, including user-tracking, ambient noise level monitoring, and mapping of auditory environments.

In summary, the auditory interface is a rapidly evolving element in human computer interaction, with a huge potential for a better use of the user's skills and perceptual resources.

5.9 Acknowledgements

We wanted to indicate the primary sources of contribution for each section. We also want to thank the reviewers for their insights and feedback on the chapter.

Introduction – S. Camille Peres

Section 1 – Virginia Best, Barbara Shinn-Cunningham, and John Neuhoff

Section 2 – Thomas Hermann

Section 3, 5 to 7 – Tony Stockman, Louise Valgeður Nickerson, and Christopher Frauenberger

Section 4 – Derek Brock and S. Camille Peres

Section 8 – Thomas Hermann and Derek Brock

5.10 References

Andersen, G. (2002). Playing by ear: Creating blind-accessible games. Gamasutra. http://www.gamasutra.com/resource_guide/20020520/andersen_01.htm.

Arons, B. (1992). A review of the cocktail party effect. *Journal of the American Voice I/O Society*, 12, 35–50.

Audyssey Magazine (1996–2006). http://www.audiogames.net/page.php?pagefile=audyssey.

Baier, G., and Hermann, T. (2004). The sonification of rhythms in human electroencephalogram. Paper presented at the International Community for Auditory Display, Syndey, Australia.

Baier, G., Hermann, T., Sahle, S., and Ritter, H. (2006). Sonified epileptic rhythms. In Proceedings of the International Conference on Auditory Display, London, UK.

Ballas, J. A. (1993). Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology: Human Perception and Performance*. 19, 250-267.

Ballas, J. A. (1994). Delivery of information through sound. In G. Kramer (Ed.), *Auditory display: Sonification, audification and auditory interfaces*. (pp. 79-94). Reading, MA: Addison-Wesley.

Barras, S. (1997). Auditory information design. Unpublished Dissertation, The Australian National University.

Bay Advanced Technologies (2006). BAT K-Sonar. Date retrieved: December 14, 2006. http://www.batforblind.co.nz/.

Berman, L. and Gallagher, K. (2006). Listening to program slices. In Proceedings of the International Conference on Auditory Display, London, UK.

Best, V., Gallun, F. J., Carlile, S. and Shinn-Cunningham, B. G. (2007). Binaural interference and auditory grouping. *Journal of the Acoustical Society of America*. 121, 1070-1076.

Blattner, M. M., Sumikawa, D. A., and Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human-Computer Interaction*, 4(1), 11–44.

Blattner, M., Papp, A., and Glinert, E. P. (1994). Sonic Enhancement of Two-Dimensional Graphics Displays. In G. Kramer (Ed.), *Auditory Display: Sonification, Audification, and Auditory Interfaces*. (pp. 447-470). Reading, MA: Addison-Wesley.

Bly, S. (1982). Presenting information in sound. Paper presented at the SIGCHI conference on Human Factors in Computing Systems.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of sound*. The MIT Press, Cambridge, Massachusetts, USA.

Brewster, S. A. (1994). Providing a structured method for integrating non-speech audio into human-computer interfaces. PhD thesis, University of York, UK. http://www.dcs.gla.ac.uk/~stephen/papers/Brewster_thesis.pdf.

Brewster, S. A. (2002). Nonspeech auditory output. In The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications archive (pp. 220 - 239). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Broadbent, D. E. (1958). *Perception and Communication*. Oxford University Press, New York.

Brock, D., Ballas, J. A., Stroup, J. L., and McClimens, B. (2004). The design of mixed-use, virtual auditory displays: Recent findings with a dual-task paradigm. Proceedings of the International Conference on Auditory Display. Sydney, Australia.

Bronkhorst, A. W. and Houtgast, T (1999). Auditory distance perception in rooms. *Nature* 397, 517-520.

Bronkhorst, A.W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. Acustica 86, 117-128.

Brown, M. L., Newsome, S. L., and Glinert, E. P. (1989). An experiment into the use of auditory cues to reduce visual workload. Paper presented at the SIGCHI conference on Human Factors in Computing Systems.

Calvert, G. A., Spence, C., and Stein, B. E. (Eds.). (2004). *The Handbook of Multisensory Processes*. Cambridge, MA: MIT Press.

Carlile, S. (1996) The physical and psychophysical basis of sound localization. In: *Virtual Auditory Space: Generation and Applications*. Ed: S. Carlile. Landes: Austin, TX,

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Cook, P.R. (2002). *Real Sound Synthesis for Interactive Applications.* A K Peters Ltd.

Darwin, C. J. and Hukin, R. W. (2000). Effectiveness of spatial cues, prosody and talker characteristics in selective attention. Journal of the Acoustical Society of America, 107, 970-977

de Campo, A. (2007). Toward a sonification design space map. Proceedings of the International Conference on Auditory Display, Montreal, Canada.

Deutsch, D. (Ed.). (1999). *The Psychology of Music*, 2nd Ed. San Diego: Academic Press.

Dix, A., Finlay, J., Abowd, G. D., and Beale, R. (2004). *Human-Computer Interaction*. Prentice Hall Europe, 3rd edition.

Etter, R. and Specht, M. (2005). Melodious walkabout: Implicit navigation with contextualized personal audio contents. In Adjunct Proceedings of the Third International Conference on Pervasive Computing. Munich, Germany.

Farkas,W. (2006). iSIC - data modeling through music. Date retrieved: December 16, 2006. http://www.vagueterrain.net/content/archives/journal03/farkas01.html.

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. (1996). *Advances in Knowledge Discovery and Data Mining*. Cambridge, MA: AAAI/MIT Press.

FindSounds. 2006, from http://www.findsounds.com/

Fitch, W. T., & Kramer, G. (1994). Sonifying the body electric: Superiority of an auditory display over a visual display in a complex, multivariate system. In G. Kramer (Ed.), *Auditory display: Sonification, audification and auditory interfaces* (pp. 307-326). Reading, MA, USA: Addison-Wesley.

Flowers, J. H., Whitwer, L. E., Grafel, D. C., and Kotan, C. A. (2001). Sonification of daily weather records: Issues of perception, attention and memory in design choices. Proceedings of the International Conference on Auditory Display, Espoo, Finland, 222-226.

Frauenberger, C., Höldrich, R., and de Campo, A. (2004). A generic, semantically based design approach for spatial auditory computer displays. Paper presented at the International Conference on Auditory Display, Sydney, Australia.

Freedom Scientific (2005). Jaws R for windows. http://www.freedomscientific.com/fs_products/JAWS_HQ.asp.

Gaver, W. W. (1994). Using and creating auditory icons. In G. Kramer (Ed.), *Auditory display: Sonification, audification and auditory interfaces*. (pp. 417–446). Reading, MA: Addison-Wesley.

Goffin, V., Allauzen, C., Bocchieri, E., Hakkani-Tür, D., Ljolje, A., Parthasarathy, S., Rahim, M., Riccardi, G., and Saraclar, M. (2005). The AT&T WATSON speech recognizer. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Philadelphia, PA, USA.

Guillaume, A., Pellieux, L., Chastres, V., and Drake, C. (2003). Judging the urgency of nonvocal auditory warning signals: Perceptual and cognitive processes. Journal of Experimental Psychology, 9(3), pp. 196-212, 2003.

GW Micro (2006). Window-eyes. http://www.gwmicro.com/Window-Eyes/.
Habrias, H. and Frappier, M. (2006). Software Specification Methods: An Overview Using a Case Study. ISTE Publishing: London.
Hart, S. G. and Staveland, L. E. (1988). Development of NASATLX (Task Load Index): Results of empirical and theoretical research. In Human mental workload (pp. 139-183). North-Holland, Amsterdam.
Hayward, C. (1994). Listening to the earth sing. In G. Kramer (ed.), Auditory display: Sonification, audification and auditory interfaces., 369-404. Reading, MA: Addison-Wesley.
Hermann, T. (2002). Sonification for exploratory data analysis. PhD thesis, Bielefeld University, Bielefeld, Germany.
Hermann, T. and Hunt, A. (2005). An introduction to interactive sonification. IEEE Multimedia.
Hermann, T., Baier, G., Stephani, U., and Ritter, H. (2006). Vocal sonification of pathologic EEG features. In Proceedings of the International Conference on Auditory Display. 158–163
Hermann, T., Drees, J. M., and Ritter, H. (2005) Broadcasting Auditory Weather Reports -- A Pilot Project, In Proceedings of the International Conference on Auditory Display, Boston, MA, USA,.
Hix, D. and Hartson, H. R. (1993). Formative Evaluation: Ensuring Usability in User Interfaces. In User Interface Software (Vol. 1, pp. 1-30). New York, NY, USA: Wiley.
Hunt, A. and Hermann, T. (2004). The importance of interaction in sonification. Paper presented at the International Conference on Auditory Displays, Sydney, Australia.
ICAD. (2007). International Community for Auditory Display. Date retrieved: July 21, 2007. http://www.icad.org/
Janata, P., and Childs, E. (2004). Marketbuzz: Sonification of real-time financial data. Paper presented at the International Conference on Auditory Displays, Sydney, Australia.
Jones, D. G. and Endsley, M. R. (2000). Overcoming representational errors in complex environments. Human Factors. 42(3), 367-378.
Kidd, G. Jr., Mason, C. R., and Arbogast, T. L. (2002). Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns. Journal of the Acoustical Society of America 111(3), 1367-1376.
Kramer, G. (1994). Auditory Display: Sonification, Audification and Auditory Interfaces. Reading, Mass.: Addison-Wesley.
Leplatre, G., and McGregor, I. (2004). How to tackle auditory interface aesthetics? Discussion and case study. Paper presented at the International Conference on Auditory Display, Sydney, Australia.
Lieder, C. N. (2004). Digital Audio Workstation. McGraw-Hill, New York.
Lumsden, J., Brewster, S. A., Crease, M., & Gray, P. D. (2002). Guidelines for Audio-Enhancement of Graphical User Interface Widgets. Paper presented at the People and Computers XVI – Memorable Yet Invisible: Human Computer Interaction, London, UK.
Mankoff, J., Dey, A. K., Hsieh, G., Kientz, J., Ames, M., and Lederer, S. (2003). Heuristic evaluation of ambient displays. Paper presented at the SIGCHI conference on Human Factors in Computing Systems-CHI Letters.
Martinson, E. and Brock D. (2007). Improving human-robot interaction through adaptation to the auditory scene. In Proceedings of the International Conference on Human-Robot Interaction. Washington, DC.
Massof, R. W. (2003). Auditory assistive devices for the blind. Proceedings of the Third International Conference on Auditory Display, Boston, MA, USA, 271-275.
McAdams, S. and Bigand E. (Eds.). (1993). *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford University Press.
McCartney, J. (1996). SuperCollider Hub. Retrieved February, 2007, from http://supercollider.sourceforge.net

McFarlane, D. C. and Latorella, K. A. (2002). The scope and importance of human interruption in HCI design. *Human-Computer Interaction*. 17(1), 1-61.

McGookin, D. K., and Brewster, S. A. (2006a). Haptic and Audio Interaction Design Paper presented at the Lecture Notes in Computer Science Springer-Verlag.

McGookin, D. K., and Brewster, S. A. (2006b). Advantage and issues with concurrent audio presentation as part of an auditory display. Paper presented at the International Community for Auditory Display, London, UK.

Mills, A. W. (1958). On the minimum audible angle. *Journal of the Acoustical Society of America* 30(4), 237-246.

Miranda, E. R. (1998). *Computer Sound Synthesis for the Electronic Musician*. Focal Press, Oxford.

Mitsopoulos, E. N. (2000). A Principled Approach to the Design of Auditory Interaction in the Non-Visual User Interface. Unpublished Dissertation, The University of York.

Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing* (5th ed.). London, UK: Academic Press.

Moore, F. R. (1990). *Elements of Computer Music*. Prentice Hall.

Mossberg, W. (2006, December 21). New earphone devices let you go cordless on iPods, cellphones. Personal Technology, The Wall Street Journal Online. Retrieved March 8, 2007, from http://ptech.wsj.com/archive/ptech-20061221.html

Mynatt, E. (1995). Transforming Graphical Interfaces into Auditory Interfaces. Unpublished Dissertation, Georgia Institute of Technology, Atlanta, GA, USA.

Neisser, U. (1967). *Cognitive Psychology*. Appleton-Century-Crofts, New York.

Neuhoff, J. (2004). Interacting auditory dimensions. In J. Neuhoff (Ed.), Ecological Psychoacoustics: Elsevier Academic Press.

Nielsen, J. (1994). Ten Usability Heuristics [Online]. Retrieved January 27, 2007, from ttp://www.useit.com/papers/heuristic/heuristic_list.htm

Nielsen, J., and Molich, R. (1990). Heuristic evaluation of user interfaces. Paper presented at the SIGCHI conference on Human Factors in Computing Systems.

Peres, S. C. and Lane, D. M. (2005). Auditory graphs: the effects of redundant dimensions and divided attention. ICAD, Limerick, Ireland.

Petrie, H., and Morley, S. (1998). The use of non-speech sounds in non-visual interfaces to the MS-windows GUI for blind computer users. Paper presented at the International Conference on Auditory Displays, Glasgow, UK.

Pirhonen, A., Murphy, E., McAllister, G., and Yu, W. (2006). Non-speech sounds as elements of a use scenario: a semiotic perspective. Paper presented at the International Community for Auditory Display, London, UK.

Plomp, R. (1964). Rate of decay of auditory sensation. *Journal of the Acoustical Society of America* 36, 277-282.

Pollack, I., and Ficks, L. (1954). Information of elementary multidimensional auditory displays. *Journal of the Acoustical Society of America,* 26, 155-158.

Puckette, M. S. (1997). Pure data. In: Proceedings of the International Computer Music Conference, 224-227.

Pulkki, V. (1997). Virtual Sound Source Positioning Using Vector Base Amplitude Panning. Journal of the Audio Engineering Society, 45(6).

Recarte, M. A. and Nunes, L. M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology, Applied*, 9(2), 119–137.

Roads, C. (2001). *Microsound*. The MIT Press, Cambridge.

Sawhney, N. and Schmandt, C. (2003). http://web.media.mit.edu/~nitin/NomadicRadio/.

Schmandt, C. (1993). Phoneshell: the telephone as computer terminal. Paper presented at the ACM International Conference on Multimedia.

Shannon, R. V., Zeng, F., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science* 270(5234), 303-304.

Shinn-Cunningham, B. G. (2005). Influences of spatial cues on grouping and understanding sound. Proceedings of the Forum Acusticum.

Staal, M. A. (2004). Stress, cognition, and human performance: a literature review and conceptual framework. *National Aeronautics and Space Administration (NASA) Technical Memorandum*, TM-2004-212824. Ames Research Center.

Stanton, N. A. and Edworthy, J., Eds. (1999). *Human Factors in Auditory Warnings*. Ashgate Publishers: Aldershot.

Stevens, S. S. (1957). On the psychophysical law. Psychological Review 64(3), 153-181.

tenHave, P. (1999). *Doing conversation analysis: a practical guide*. Sage, London, UK.

Terenzi, F. (1988). Design and realization of an integrated system for the composition of musical scores and for the numerical synthesis of sound (special application for translation of radiation from galaxies into sound using computer music procedures). Physics Department-University of Milan.

Truax, B. (Ed.). (1999). *Handbook for Acoustic Ecology*. Date retrieved: July 21, 2007. http://www2.sfu.ca/sonic-studio/handbook/.

Ulfvengren, P. (2003). *Design of Natural Warning Sounds in Human Machine Systems.* Doctoral thesis, Royal Institute of Technology Stockholm, Sweden.

VRSonic (2007). VibeStudio. Date retrieved: July 20, 2007. http://www.vrsonic.com.

Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. *Journal of Experimental Psychology: Applied*, 8(4), 211-221.

Walker, B. N. and Kramer, G. (2004). Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making. In J. Neuhoff (Ed.), *Ecological psychoacoustics (*pp.150-175). New York: Academic Press.

Walker, B. N. and Lane, D. M. (2001). Psychophysical scaling of sonification mappings: A comparison of visually impaired and sighted listeners. Proceedings of the Seventh International Conference on Auditory Display, Espoo, Finland, 90-94.

Walker, B. N. and Lindsay, J. (2004). Auditory navigation performance is affected by waypoint capture radius. Paper presented at the International Conference on Auditory Display, Sydney, Australia.

Walker, B. N. and Stamper, K. (2005). Building audio designs monkey: An audio augmented reality designer's tool. Paper presented at the International Conference on Auditory Display, Limerick, Ireland.

Walker, B. N., Nancy, A., and Lindsay, J. (2006). Spearcons: speech-based earcons improve navigation performance in auditory menus. In Proceedings of the International Conference on Auditory Display, London, UK.

Wang, D. and Brown, G. J. (2006). *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*. Hoboken, NJ: John Wiley & Sons.

Weir, J. (2005). The Sound of Traffic. Date retrieved: December 16 2006. http://www.smokinggun.com/projects/soundoftraffic/.

Wickens, C. D. and Hollands, J. G. (2000). *Engineering Psychology and Human Performance*, 3rd Ed. Prentice Hall.

Williamson, J., Murray-Smith, R., and Hughes, S. (2007). Shoogle: Multimodal Excitatory Interfaces on Mobile Devices. Paper presented at the SIGCHI conference on Human Factors in Computing Systems.

Winberg, F. and Bowers, J. (2004). Assembling the senses: towards the design of cooperative interfaces for visually impaired users. Paper presented at the ACM conference on Computer supported cooperative work.